


Predicting death risk analysis in fully vaccinated people using novel extreme regression-voting classifier

Digital Health
Volume 8: 1–20
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076221109530
journals.sagepub.com/home/dhj



Eysha Saad^{1,*}, Saima Sadiq^{1,*}, Ramish Jamil¹, Furqan Rustam²,
Arif Mehmood³, Gyu Sang Choi⁴ and Imran Ashraf⁴ 

Abstract

Vaccination for the COVID-19 pandemic has raised serious concerns among the public and various rumours are spread regarding the resulting illness, adverse reactions, and death. Such rumours can damage the campaign against the COVID-19 and should be dealt with accordingly. One prospective solution is to use machine learning-based models to predict the death risk for vaccinated people by utilizing the available data. This study focuses on the prognosis of three significant events including ‘not survived’, ‘recovered’, and ‘not recovered’ based on the adverse events followed by the second dose of the COVID-19 vaccine. Extensive experiments are performed to analyse the efficacy of the proposed Extreme Regression-Voting Classifier model in comparison with machine learning models with Term Frequency-Inverse Document Frequency, Bag of Words, and Global Vectors, and deep learning models like Convolutional Neural Network, Long Short Term Memory, and Bidirectional Long Short Term Memory. Experiments are carried out on the original, as well as, a balanced dataset using Synthetic Minority Oversampling Approach. Results reveal that the proposed voting classifier in combination with TF-IDF outperforms with a 0.85 accuracy score on the SMOTE-balanced dataset. In line with this, the validation of the proposed voting classifier on binary classification shows state-of-the-art results with a 0.98 accuracy.

Keywords

COVID-19, post-vaccination symptoms, adverse reactions, machine learning

Submission date: 20 May 2022; Acceptance date: 8 June 2022

Introduction

The last two decades have witnessed many pandemics like SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome), COVID-19 (coronavirus disease 2019), etc. Recently, COVID-19 infected approximately 308 million people in 223 countries leading to 5.492 million deaths as of 12 January 2020¹. The ongoing COVID-19 pandemic impacted the individual, as well as, the public life of human beings on a global scale, and containing it seems to be very difficult in the near future. Although, it possibly can be confined like other viruses, such as HKU1, NL63, 229E, and OC43, however, the substantial human and financial loss remains the main concern². Precautionary measures against COVID-19, such as sanitation procedures, physical

¹Department of Computer Science, Khawaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan

²Department of Software Engineering, University of Management and Technology, Lahore, Pakistan

³Department of Computer Science & Information Technology, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

⁴Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, Republic of Korea

*The authors Eysha Saad and Saima Sadiq contributed equally to this work.

Corresponding authors:

Gyu Sang Choi, Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea.
Email: castchoi@ynu.ac.kr

Imran Ashraf, Department of Information and Communication Engineering, Yeungnam University, Gyeongsan, 38544, Republic of Korea.
Email: imranashraf@ynu.ac.kr



distancing, personal hygiene, mask usage, disinfection of the surfaces, and frequent hand washing are essential to reduce its spread. However, the case fatality ratio (CFR), a measure of mortality among infected cases, continues to increase³. Facilitating a safe return to normal life along with minimization of the COVID-19 resurgence requires the immunity against COVID-19⁴ which is aimed by several developed vaccines like Moderna, Pfizer (BioNTech), and Johnson & Johnson, etc.⁵ As of December 2020, several vaccines have been administered with different efficiency and immunity against COVID-19, as shown in Figure 1.

Similar to vaccines for other diseases, COVID-19 vaccines have been reported for several side effects. Reports of adverse side effects following the doses of COVID-19 vaccination are submitted to VAERS (Vaccine Adverse Event Reporting System). From 1 January 2021 to 19 March 2021, a total of 5351 adverse events have been reported to VAERS. The adverse side effects range from mild to severe such as fever, pain, diarrhoea, fatigue, blood pressure, chills, muscle pain, headache, and pain at the injection site and are shown in Figure 2(a). Similarly, several COVID-19 positive cases are reported after being vaccinated. Further include dizziness and severe allergic reactions. Blood clotting, cardiac problems, and resulting deaths are also reported following adverse events such as cardiac arrest, abdominal pain, etc. as shown in Figure 2(b). There is also a theoretical risk that vaccination could make infection severe by enhancing the respiratory disease⁶. Such adverse reaction and death reports make it significantly important to analyse the data regarding the adverse effects of COVID-19 vaccines and report reactions with a higher probability of fatality to assist healthcare professionals in prioritizing the cases with adverse effects and provide timely medical treatment.

The ML (machine learning) is the self-regulated discovery of potentially valid or useful knowledge and novel hidden patterns from dataset⁷. ML models operate by revealing relationships and patterns among the data instances in single or multiple datasets. ML has been widely applied in the healthcare sectors for its applications in simulating health outcomes, forecasting patient outcomes, and evaluating medicines⁸. In recent years, ML has also been extensively used in the diagnosis and prognosis of many diseases like COVID-19, as immense data is being generated regarding COVID-19 on an everyday basis, which can be analysed to predict the COVID-19 case and devise corresponding policies to contain the pandemic. In the same vein, data associated with adverse events reports post-COVID-19, gathered by VAERS was made public on 27 January 2021 which motivated current research.

This study demonstrates an enhanced ML-based prediction system to analyse the adverse events associated with the COVID-19 vaccine and predict individuals with

symptoms that might cause fatality so that healthcare professionals can treat the individuals beforehand. It helps medical experts critically monitor vaccinated individuals with death risks. This study makes the following major contributions:

- This study advocates a systematic approach to investigate the adverse events following the COVID-19 vaccine for possible death leading symptoms. The prognosis of three significant events including ‘not survived’, ‘recovered’, and ‘not recovered’ is made in this regard.
- A novel vote-based ER-VC (Extreme Regression-Voting Classifier) is devised which combines ET and LR under soft voting criterion to increase the prediction accuracy. Extensive experiments are carried out for performance analysis concerning many machine learning models like RF (Random Forest), LR (Logistic Regression), MLP (Multilayer Perceptron), GBM (Gradient Boosting Machine), AB (AdaBoost), kNN (k Nearest Neighbours), and ET (Extra Tree Classifier). In addition LSTM (Long Short Term Memory), CNN (Convolutional Neural Network), and BiLSTM (Bidirectional LSTM) are also implemented for appraising the performance of the proposed approach.
- To analyse the influence of data balancing, the performance of ML models is analysed and compared by integrating SMOTE (Synthetic Minority Oversampling Technique) for predicting the survival of vaccinated individuals.

The structure of this research is organized into five sections. Section ‘Related work’ represents the previous works related to this study. Later, the proposed approach, ML models, and dataset description are provided in Section ‘Material and methods’. Section ‘Results and discussion’ provides the analysis and discussion of the results. In the end, the study is concluded in Section ‘Conclusion’.

Related work

The COVID-19 pandemic inflicted substantial economic and human losses worldwide. With unusual symptoms, the disease is difficult to treat based on previous methods used for treatment. However, the strong infrastructure of electronic health records and advanced technologies in recent times has helped in conducting several research studies and exploration of its treatment. The data repositories of COVID-19 patients’ symptoms and track records are maintained efficiently by medical and government institutions to explore health risks. Laboratory tests, radiological reports, and patients’ symptoms have been analysed using ML models by many researchers. Early studies, mostly focused on disease diagnoses and predicting the death rate of COVID-19 patients based on statistical models⁹. After some time, hospital records of patients are mostly used to identify potential risks¹⁰.

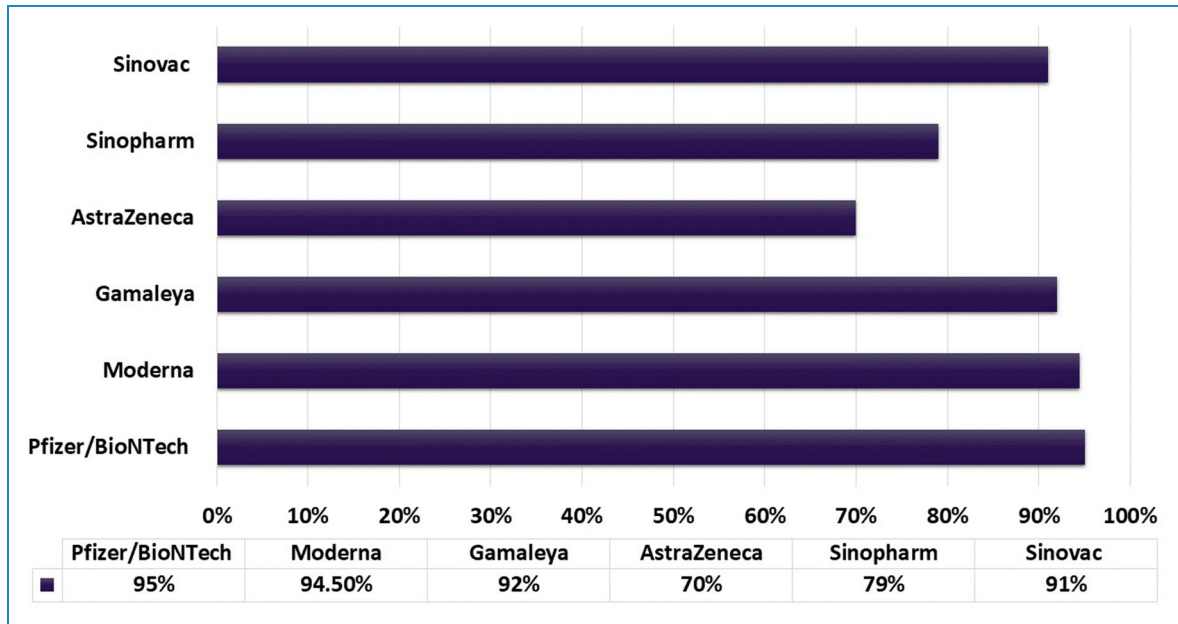


Figure 1. Efficacy of COVID-19 vaccines.

The exacerbated outbreak of the COVID-19 pandemic and its potential risk to human lives necessitated different medical research laboratories and pharma industries to start developing the COVID-19 vaccine at a fast pace. For providing herd immunity to people, there was a need for a safe and effective vaccine in a short time¹¹. At the end of 2020, 48 vaccines were available at the clinical trial phase, and three vaccines including Pfizer, Moderna, and AstraZeneca completed this phase in the US¹². During the first phase, millions of health professionals were vaccinated, then populations at higher risk, such as people older than 65 years are covered¹³.

Severe outcomes leading to the death risk of COVID-19 patients are associated with different pre-existing medical conditions and comorbidities^{14,15}. Approximately more than 40% of patients hospitalized with COVID-19 had at least one comorbidity¹⁶. In a similar study, the authors analysed comorbidities between survivor and non-survivor patients¹⁷. Common diseases included diabetes mellitus, cardiovascular disease, chronic obstructive pulmonary disease, hypertension, and kidney-related diseases. Various other biomarkers such as C-reactive protein, high level of ferritin, white lymphocyte count, blood cell count, procalcitonin, and d-dimer are related to health risks and are increasing the mortality rate of COVID-19 patients¹⁸. These biomarkers and other symptoms could offer advantages in predicting death risks.

Various types of deep learning architectures have also been employed for different tasks. For example, the bidirectional neural network is proposed by Onan¹⁹ that uses a group-wise enhancement mechanism for feature extraction. By dividing features into multiple groups, important

features from each group can be obtained to increase the performance. Similarly, a bidirectional LSTM model is presented by Onan and Korukoğlu²¹ that combines term weighting using inverse gravity moment with trigrams. Ensemble models are also reported to produce better results for sentiment analysis tasks^{22,23}. Such models utilize different ensemble schemes, clustering, and feature extraction approaches for increased performance. For example, Onan²⁴ devises a feature extraction approach for sentiment analysis while Onan et al.²⁵ follows a hybrid ensemble model using the concept of consensus clustering. Similarly, Onan^{26,27} adopts ensemble models for sentiment analysis and opinion mining²⁸. Along the same lines, topic modelling is focused on using ensemble models by Onan^{29,30}. The topic of sarcasm detection is covered by Onan³¹ by following a hybrid model approach while Sadiq et al.³² investigates aggression detection. The authors have explored many ML-based techniques using patients' symptoms and laboratory reports during hospitalization³³. Researchers are diligent in defeating COVID-19 by exploring ways of COVID-19 detection³⁴ and devising frameworks to control the spread of disease³⁵. Researchers applied an ML model to electronic health records to predict the mortality rate of COVID-19 patients³⁶. However, the non-infected population is getting benefits from vaccination. Because of heterogeneity among the population due to demographic categories, risk patterns regarding COVID-19 disease and vaccine are difficult to predict. Different factors are involved in predicting death risks such as unique health history, obesity, cancer history, hereditary diseases, and different immunity levels. Medical professionals are striving to allocate

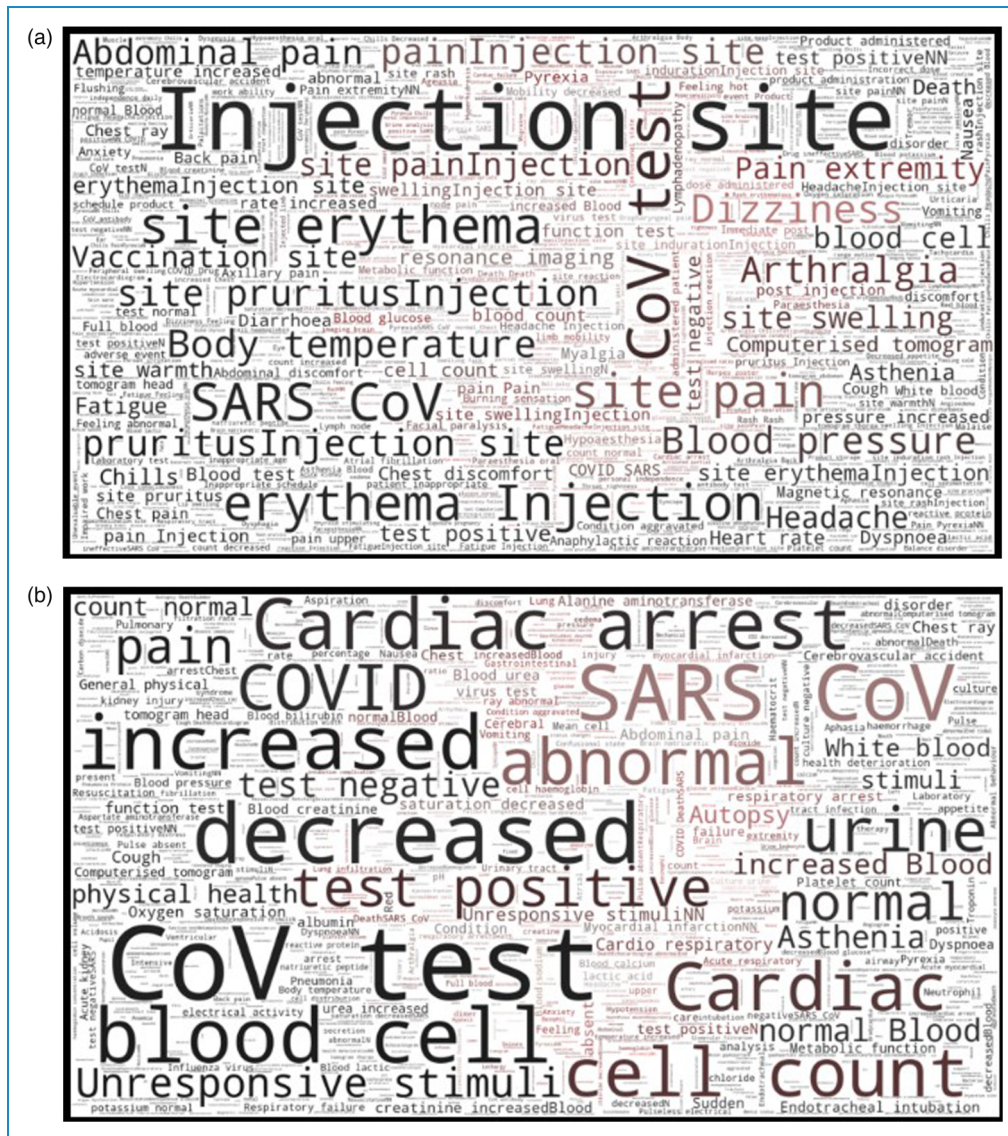


Figure 2. Word cloud of reported adverse reactions, (a) side effects following the doses of COVID-19 based on reports submitted to Vaccine Adverse Event Reporting System (VAERS), and (b) side effects of death cases post COVID-19 vaccine.

resources and provide help in maximizing the survival probability.

This study makes a significant contribution toward maximizing the survival rate of vaccinated people by predicting the probability of fatal outcomes beforehand by analysing the post-vaccination symptoms. We leveraged growing electronic records and advanced predictive analytical methods to predict the risk associated with the side effects of COVID-19 vaccines.

Material and methods

This study works on the highly accurate prognosis of death risk patients in addition to recovered and not recovered cases concerning the adverse events reported after the

second dose of the COVID-19 vaccine. Experiments in this research can be categorized into two stages where Stage I deals with the multiclass classification of adverse events as ‘not survived’, ‘recovered’, and ‘not recovered’ while Stage II or validation stage is concerned with the binary classification of the adverse reactions into ‘survived’ and ‘not survived’. This section contains a brief description of the dataset utilized in this study, as well as, the proposed methodology adopted for classification tasks.

Dataset description

This study utilizes the COVID-19 VAERS dataset acquired from Kaggle which is an open repository for benchmark datasets³⁷. The dataset contains the adverse events reported

Table 1. Description of data attributes of COVID-19 World Vaccine Adverse Reactions dataset.

Variable	Description
VAERS_ID	Identification number for each vaccinated case
RECVDATE	Receiving date of adverse reactions report
STATE	Region of the country from which report was received
AGE_YRS	Age of vaccinated individual
CAGE_YR	Age calculation of individual in years
CAGE_MO	Age calculation of vaccinated individual in months
SEX	Gender of vaccinated individual
RPT_DATE	Date on which report form was completed
SYMPTOM_TEXT	Reported symptoms
DIED	Survival status
DATEDIED	Date of death of vaccinated individual
L_THREAT	Severe illness
ER_VISIT	Visited doctor or emergency room
HOSPITAL	Is hospitalized or not
HOSPDAYS	Number of days individual was hospitalized
X_STAY	Elongation of hospitalized days
DISABLE	Disability status of vaccinated individual
RECOVD	Recovery status of vaccinated individual
VAX_DATE	Date on which individual was vaccinated
ONSET_DATE	Onset date of adverse event
NUMDAYS	ONSET_DATE-VAX_DATE
LAB_DATA	Laboratory reports
V_ADMINBY	Vaccine administration facility
V_FUNDBY	Funds used by administration to buy vaccine
OTHER_MEDS	Other medicines in use by vaccinated individual

(continued)

Table 1. Continued.

Variable	Description
CUR_ILL	Information regarding illness of individual at the time of getting vaccinated
HISTORY	Long-standing or chronic health-related conditions
PRIOR_VAX	Information regarding prior vaccination
SPLTTYPE	Manufacturer Report Number
FORM_VERS	Version 1 or 2 of VAERS form
TODAYS_DATE	Form completion date
BIRTH_DEFECT	Birth defect
OFC_VISIT	Clinic visit
ER_ED_VISIT	Emergency room visit
ALLERGIES	Allergies to any product

by individuals after the COVID-19 vaccine along with details related to the particular individuals³⁸. It consists of a total of 5351 records and 35 variables, details of which are given in Table 1. The study is concerned with investigating the death risk of vaccinated individuals by analysing the adverse events. On that account, we utilized only three variables such as ‘RECOVD’, ‘DIED’, and ‘SYMPTOM_TEXT’ for multiclass classification and two variables including ‘DIED’ and ‘SYMPTOM_TEXT’ for binary class classification. The variable ‘DIED’ comprises two classes involving ‘survived’ and ‘not survived’ corresponding to 4541 and 810 records, respectively. Whereas, the variable ‘RECOVD’ comprises three target variables, including ‘recovered’, ‘not recovered’, and ‘recovery status unknown’ corresponding to 1143, 2398, and 1810 records, respectively. Some of the ‘DIED’ cases are regarded as ‘not recovered’ while some belong to the ‘recovery status unknown’ category as shown in Figure 3(a). The correspondence between the ‘DIED’ and ‘RECOVD’ features shows that a portion of the cases which did not recover from COVID-19 did not survive after being vaccinated. Figure 3(b) reveals that adverse events leading to the death of the vaccinated individuals comprise 15% of the dataset which shows that there is an unequal distribution of class in both binary class and multiclass distribution. For an effective analysis, we disregarded the records which correspond to ‘recovery status unknown’ except for the ones which belong to the ‘not survived’ category in the multiclass classification.

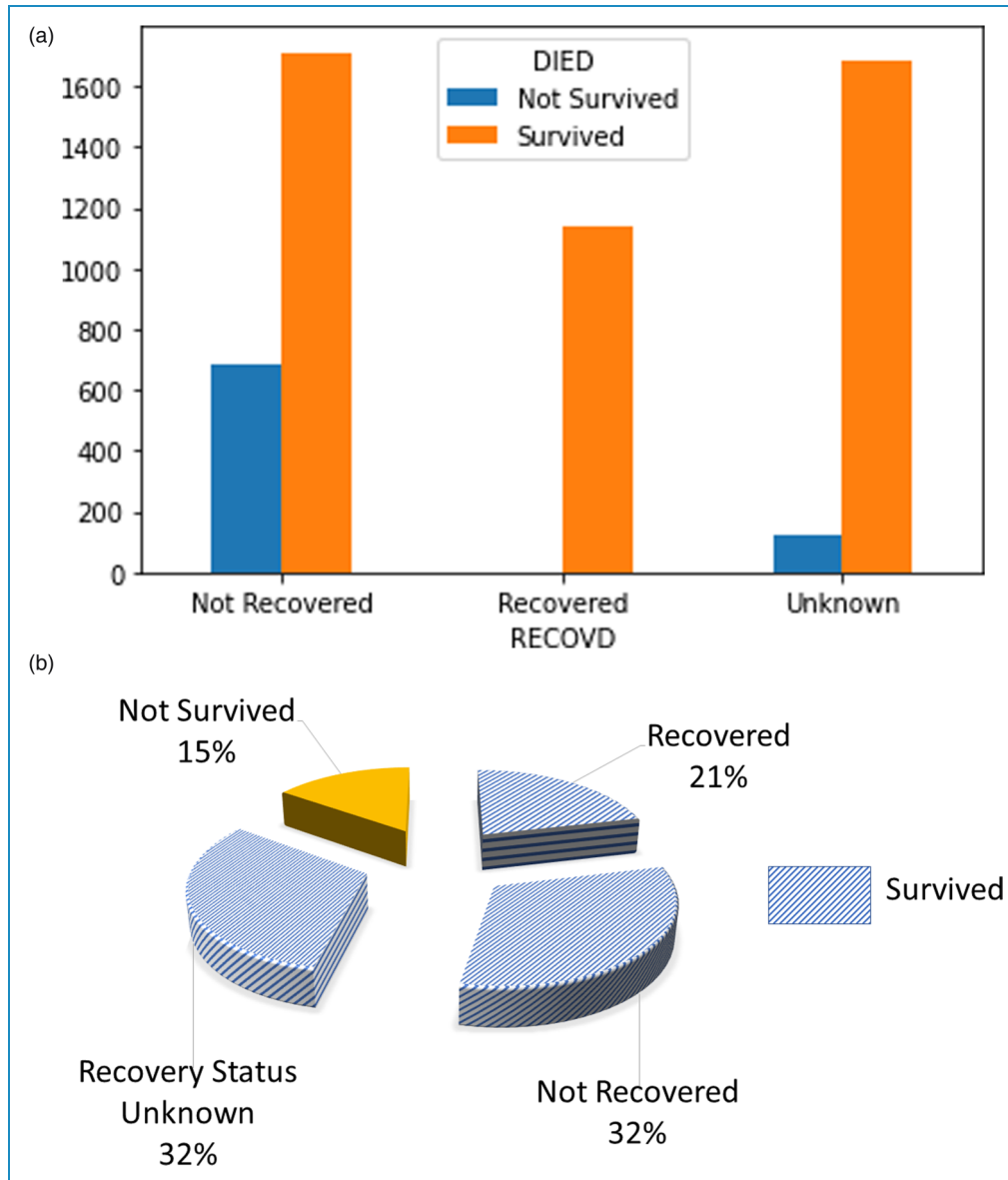


Figure 3. Dataset visualization, (a) correspondence between the categories related to 'DIED' and 'RECOVD' features, and (b) class distribution.

Problem statement

Consider an individual X who has received his second dosage of the COVID-19 vaccine. Although there are some minor side effects of the COVID-19 vaccine in some cases, the side effects can cause death. To maximize the survival rate and to notify healthcare professionals beforehand, our study mines the adverse reactions of the COVID-19 vaccine reported to VAERS, for the prognosis of death risks. This research looks at two scenarios: multiclass classification for recovery and survival rate analysis, and binary classification for survival rate analysis. The multiclass categorization is intended to help healthcare professionals evaluate the recovery status of vaccinated

individuals as well as their fatality status and it deals with four classes. In contrast, we devised the binary classification for emergency circumstances so that patients with a significantly higher risk can be treated ahead of time. It provides the models' accuracy regarding the prediction of survival chances of vaccinated people and helps health professionals to treat the people at risk accordingly.

Proposed methodology

In this study, ML-based techniques are utilized for the analysis of adverse events caused by the COVID-19 vaccine.

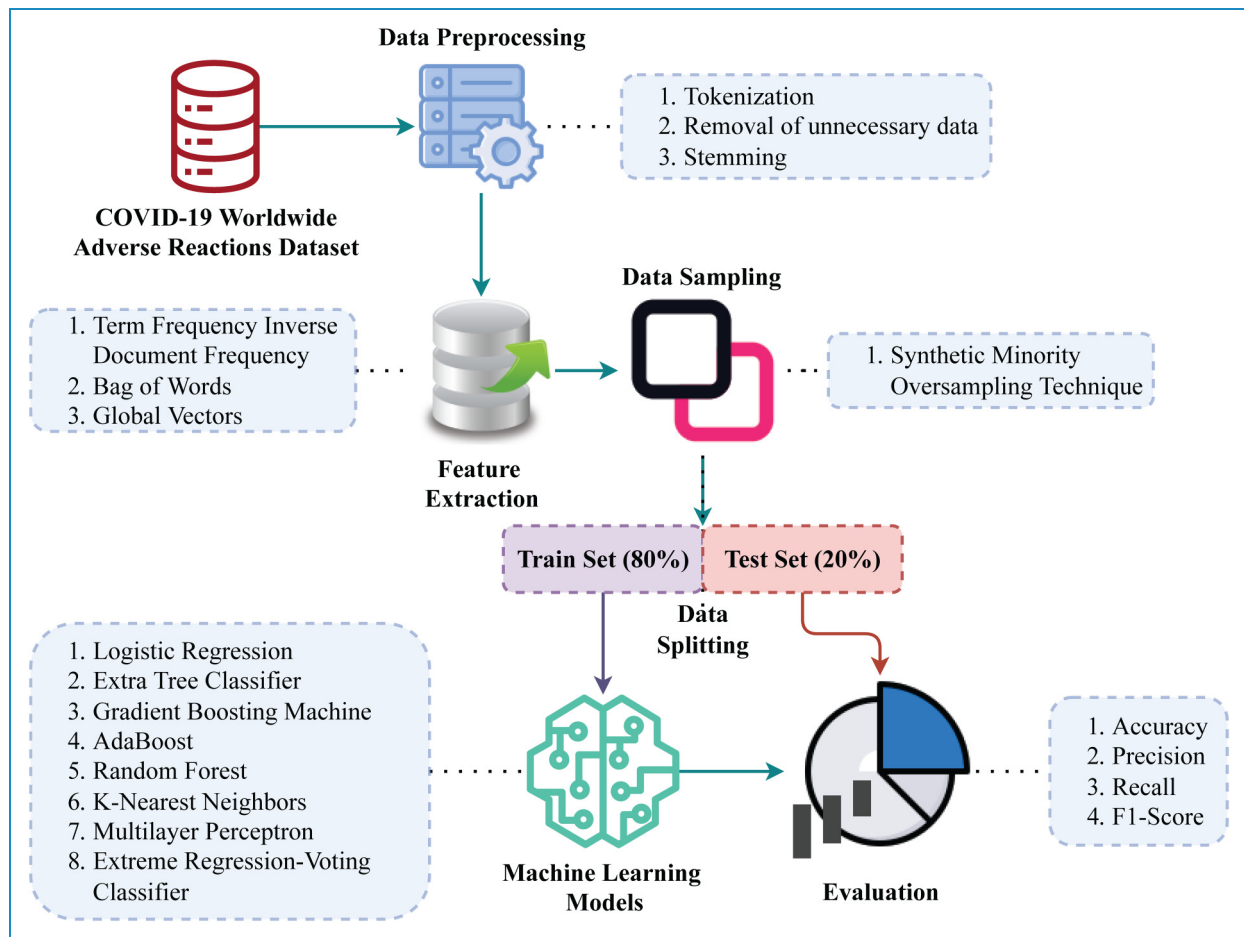


Figure 4. Architecture of the methodology devised for prognosis of death risks.

Table 2. Data count after implementation of synthetic minority oversampling approach (SMOTE) in accordance with each target variable.

Target variables	Multiclass classification		Binary classification	
	Original	SMOTE	Original	SMOTE
Survived	-	-	4541	4541
Not survived	810	1712	810	4541
Recovered	1142	1712	-	-
Not recovered	1712	1712	-	-
Total records	3664	5136	5351	9028

Figure 4 shows an architecture of the methodology adopted for the diverse range of experiments which is followed by each prediction model.

This study mainly follows multiclass classification which involves classifying adverse reactions as ‘not-survived: vaccinated individuals that died due to adverse reactions’, ‘recovered: vaccinated individuals that recovered from COVID-19’, and ‘not recovered: individuals that were tested positive of COVID-19 after vaccination’. In line with this, we integrated two data attributes including ‘RECOVD’, and ‘DIED’ as the target class, and one attribute ‘SYMPTOM_TEXT’ as a feature set in our experiments. The ‘RECOVD’ data attribute has three values including Y (recovered), N (not recovered), and U (recovery status unknown). We only utilized Y and N values from ‘RECOVD’ and Y values from ‘DIED’ for stage I experiments. This resulted in a total of 4351 instances out of which 810 instances correspond to the ‘not survived’ target variable, 2398 as ‘not recovered’, and 1143 instances are labelled as ‘recovered’. This shows the uneven distribution of target variables that can substantially dissipate the performance of classifiers. To overcome this problem, we oversampled the minority target variable using SMOTE. Oversampling by SMOTE for binary and multiclass classification is summed up in Table 2.

To reduce the training and generalize the learning patterns for the classifiers, we integrated two feature extraction techniques including BoW (Bag of Words), TF-IDF (Term Frequency-Inverse Document Frequency), and GloVe (Global Vectors). Afterwards, data is split into train and test sets with a ratio of 0.8–0.2. The number of train and test records corresponding to multiclass and binary classification is given in Table 3. Furthermore, ML classifiers, such as LR, ET, RF, GBM, AB, KNN, MLP, and proposed voting classifier learn the patterns regarding the target variable from the train set. Trained models are then tested on the unseen test data and evaluated under the criteria of accuracy, precision, recall, and F1 score.

Data preprocessing

Data preprocessing aims at enhancing the quality of the raw input data to extract meaningful information from the input data. It is followed by the preparation of input data which includes cleaning and organization of the raw data to effectively build and train the ML-based classifiers. In the current study, various steps are taken to clean, normalize and transform the ‘SYMPTOM_TEXT’. We removed irrelevant data, including punctuation, numeric, and null values from the input data. ML classifiers are prone to case sensitivity, for their efficient training we normalized the case of text by converting the text into lowercase. Afterward, we performed stemming using PorterStemmer(), and NLTK (Natural Language Tool Kit) function, for the conversion of verbs into their root forms. As the last step of preprocessing, we removed stop words that are the most frequent in the text and are not significant for the classification.

Feature extraction

Feature extraction is a technique that involves the extraction of significant and effective features from the preprocessed data for improved performance of predictive models on the unseen data. It follows the procedure of transformation of arbitrary data and finding features that are correlated with the target variable. ML classifiers guided by feature extraction technique tend to produce more accurate results³⁹. Two

feature extraction techniques including BoW, TF-IDF, and GloVe are utilized in this study.

BoW is the vectorization of text data into numeric features. It represents the word frequency within the text regardless of the information concerning its structure or position in the text. This technique considers each word as a feature⁴⁰. It does not regard the number of times different terms appear in a document. A term’s presence in a corpus is the only factor that affects its weight.

TF-IDF quantifies a word in a document by computing the weight of each word which in turn shows the significance of a word in that text⁴¹. The weight is determined by combining two metrics, TF (Term Frequency) which is a measure of the frequency of a word in a document, and IDF (Inverse Document Frequency) which refers to the measure of the frequency of a word in the entire set of documents. Here document can be considered as ‘SYMPTOM_TEXT’ in the dataset. TF-IDF for the frequency of a word x in document y can be computed as follows:

$$TF-IDF = \frac{f_{(x,y)}}{n_y} \times \log\left(\frac{N}{df(x) + 1}\right) \quad (1)$$

where $f_{(x,y)}$ is the frequency of word x in ‘SYMPTOM_TEXT’ records (y), n_y is the total number of words occurring in the y , N is the total number of ‘SYMPTOM_TEXT’ records, and $df(x)$ is the number of ‘SYMPTOM_TEXT’ records in which word x is present.

GloVe generates word embeddings of the given ‘SYMPTOM_TEXT’ by mapping the relationship between the words. This is mainly done by aggregating the global co-occurrence matrices which provide information regarding the frequency of word pairs occurring together. Similar words are clustered together and different words are discarded based on the co-occurrence matrix of a corpus. Rather than training on the entire sparse matrix or individual context windows in a large corpus, the Glove model takes advantage of statistical information as exclusively nonzero elements in a word-word co-occurrence matrix⁴².

Data sampling

When a target variable is distributed unevenly in a dataset, it leads to a misleading performance by the ML models. The reason for this is that ML models learn the decision boundary for the majority class with more efficacy than the minority class. Therefore, showing poor performance in the prediction of minority class results in ambiguous and misleading results. Hence, changing the composition of an imbalanced dataset is one of the most well-known solutions to the problem of classifying an imbalanced dataset⁴³. It can be done in two ways: undersampling or oversampling. Undersampling randomly reduces the majority class size

Table 3. Data split count corresponding to training and test sets.

Split set	Multiclass classification		Binary classification	
	Original	SMOTE	Original	SMOTE
Train set	2931	4108	4281	7211
Test set	733	1028	1070	1817

SMOTE: synthetic minority oversampling approach.

and is mostly utilized when there is an ample amount of data instances whereas, oversampling arbitrarily duplicates the minority class and is effective when implemented on a small dataset. Since we have a limited number of records in our dataset, therefore, oversampling is the best fit for the proposed framework. One of the oversampling techniques is SMOTE⁴⁴ which is utilized in the current study.

A SMOTE selects the data samples which are relatively close in the feature vector space and draws a line between those data samples⁴⁵. It then generates synthetic data samples by finding k nearest neighbours for that particular data sample with $k = 5$. This results in simulated data samples that are comparatively at a close distance in the feature space from the data samples from the minority class.

ML classifiers

Supervised ML classifiers are utilized in this study for the prediction of target variables from the data. Implementation of ML classifiers is done in Python language using the ‘scikit learn’ module. ML classifiers are trained on data samples from the training set and tested using a test set that is unknown to the classifiers. ML classifiers integrated in this study are briefly discussed here and their corresponding hyperparameter settings are given in Table 4.

Random Forest is a tree-based ML classifier that integrates aggregated results obtained by fitting many decision trees on randomly selected training samples. Each decision tree in RF is generated based on selection indicators such as Gini Index, Gain Ratio, and Information Gain to select an attribute. It is a meta-estimator that can be used both for regression and classification tasks⁴⁶.

Table 4. Hyperparameter settings of supervised machine learning classifiers.

Model	Hyperparameter settings
RF	n_estimators=100, random_state=50, max_depth=300
AB	n_estimators=100, random_state=50
LR	random_state=50, solver='saga', multi_class='ovr', C=3.0
MLP	random_state=50, max_iter=200
GBM	n_estimators=100, learning_rate=1, random_state=50
ET	n_estimators=100, random_state=50, max_depth=300
KNN	n_neighbors=5

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier.

AdaBoost also referred to as adaptive boosting is an iterative ensemble technique and is a good choice for constructing ensemble classifiers. Combining numerous weak learners into strong learners, it generates robust results. It is trained on weighted examples and provides optimized output by minimizing the error rate at each iteration⁴⁷. AdaBoost adjusts weight with respect to the classification results at each iteration. If the classification results are correct, weights for the training samples are increased while the weights are decreased for those samples which are misclassified. AdaBoost performs better due to its diversity of expansion, that is, it contains diverse classifiers.

Extra Tree Classifier is a collection of several de-correlated decision trees built from random sets of features extracted from training data. Each tree selects the best feature by computing its Gini Importance. ET incorporates averaging to control overfitting and enhance predictive accuracy⁴⁸.

Logistic Regression is a statistical ML classifier that processes the mapping between a given set of input features and a discrete set of target variables by approximating the probability using a sigmoid function. The sigmoid function is an S-shaped curve that restricts the probabilistic value between the discrete target variables as defined in equation (3). It works efficiently for classification tasks⁴⁹.

$$\sigma(x) = e^x / (e^x + 1) \quad (2)$$

$$\sigma(x) = 1 / (1 + e^{-x}) \quad (3)$$

where $\sigma(x)$ shows the output in the range of 0 and 1, x is the input, and e is the base of the natural log.

Multilayer Perceptron is an extensive feed-forward neural network that consists of three layers-input, output, and hidden layer. MLP works by receiving input signals which need to be processed at the input layer and performing predictions at the output layer. The hidden layer is the significant computational mechanism of MLP, which is situated in the middle of the input layer and the output layer. MLP is designed to map a nonlinear relationship between input and its corresponding output vector⁵⁰.

Gradient Boosting Machine is a boosting classifier that builds an ensemble of weak learners in an additive manner which proves to be useful in enhancing the accuracy and efficiency of the learning model. It employs gradient (two or more derivatives of a similar function) to identify the error in the preceding weak learner. Each weak learner in GBM attempts to minimize the error rate of the previous weak learner. It does so by integrating loss function with the gradients. It efficiently handles the missing values in the data⁵¹.

K-nearest neighbours is a straightforward ML classifier that maps the distance between a dependent variable and a target variable by adopting a particular number of k samples adjacent to the target variable. For classification,

kNN predicts by considering the majority votes of the neighbouring data points for the prevalent target variable⁵².

Proposed extreme regression-voting classifier

ER-VC is a voting classifier that aggregates the output predictions of ET and LR to generate a final output. LR determines the significance of each feature of trained samples along with providing the direction of its association with less time consumption. This makes LR a good fit for our proposed voting classifier. Consequently, ET has been selected due to its randomizing property which restrains the model from overfitting. The foundation of the proposed classifier is building an individual strong model instead of discrete models with low accuracy results. It incorporates similar hyperparameter tuning of respected classifiers as described in Table 4. ER-VC is supported with soft voting criteria such that, it generates a final prediction by averaging the probability p given to the target class. The framework of the proposed ER-VC model is illustrated in Figure 5.

The working of the proposed ER-VC classifier is illustrated in Algorithm 1. We can compute the target class for the weights assigned to predictions $l_1, l_2, l_3, \dots, l_n$ made by classifier LR and $e_1, e_2, e_3, \dots, e_n$ by classifier ET respectively as

$$P(L) = l_1, l_2, l_3, \dots, l_n \quad (4)$$

$$P(E) = e_1, e_2, e_3, \dots, e_n \quad (5)$$

$$Final\ Prediction = \underset{i=0}{\operatorname{argmax}} \sum^n (P(L) + P(E)) \quad (6)$$

where $P(L)$ and $P(E)$ are the predictions made by LR and ET, respectively.

Evaluation criteria

When a model is proposed, it is crucial to evaluate its performance. Four outcomes are produced by ML models when tested with a test set, these outcomes include TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative). TP shows the correctly predicted positive instances, TN shows correctly predicted negative instances, FP are wrongly predicted positive instances, and FN is wrongly predicted negative instances. Using these outcomes we evaluated the efficacy of our proposed framework regarding the accuracy, precision, recall, and F1 score. Where accuracy is the measure of correctness of the model, precision is the measure of the proportion of correctly predicted positive instances, recall is the measure of correctly identified positive instances, and F1 score is the harmonic mean of precision and recall. Mathematical formulas of the aforementioned evaluation parameters are

given here

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$F1\text{-score} = \frac{2 \times recall \times precision}{recall + precision} \quad (10)$$

Results and discussion

Extensive experiments have been performed using different scenarios for the prediction of three significant events in COVID-19 vaccinated people. In each scenario, ML

Algorithm 1. Algorithm for proposed Extreme Regression-Voting Classifier (ER-VC)

Input: SYMPTOM_TEXT

Output: Vaccinated individual \leftarrow not survived or recovered or not recovered

Procedure: Data Splitting

Training_Samples = (SYMPTOM_TEXT, Labels)

Testing_Samples = (SYMPTOM_TEXT)

Training_Samples, Testing_Samples

Procedure: Voting Classifier

voting_criterion= "soft"

$L = \text{Logistic_Regression}(\text{Training_Sample})$

$E = \text{Extra_Tree}(\text{Training_Sample})$

Procedure: Predictions made by L

$P(L) \leftarrow \text{Testing_samples}$

$P(L) = l_1, l_2, l_3, \dots, l_n$

Procedure: Predictions made by E

$P(E) \leftarrow \text{Testing_samples}$

$P(E) = e_1, e_2, e_3, \dots, e_n$

$Final_Prediction \leftarrow \underset{i=0}{\operatorname{argmax}} \sum^n (P(L) + P(E))$

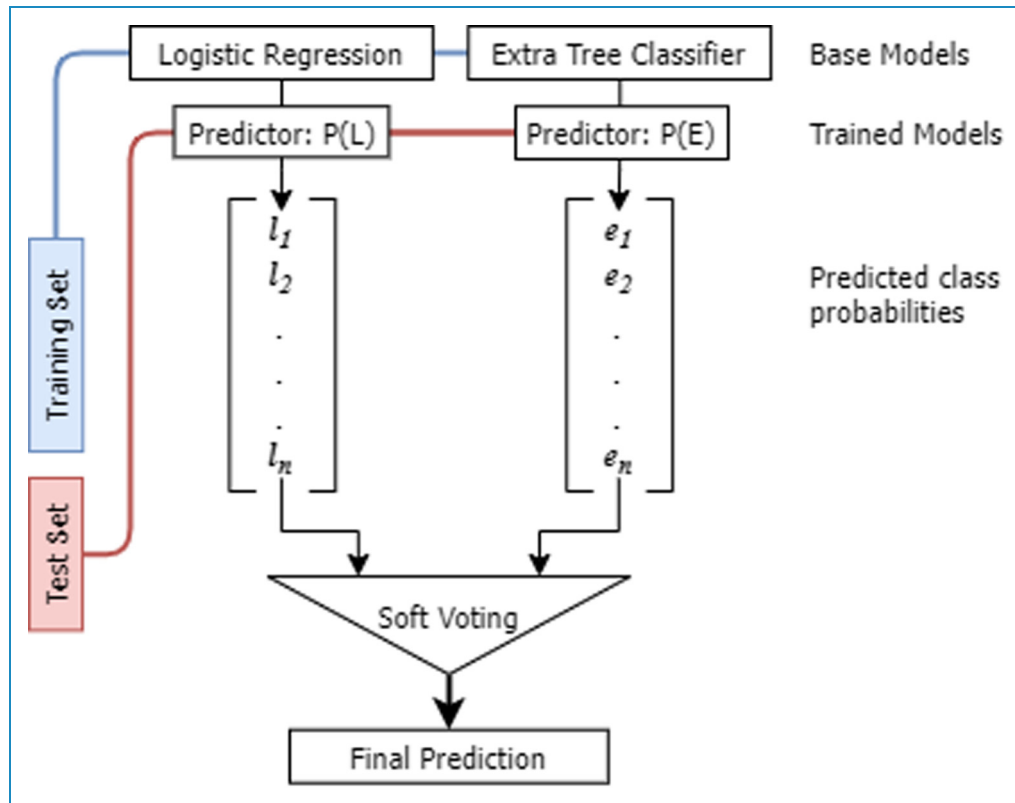


Figure 5. Framework of extreme regression-voting classifier.

Table 5. Classification results of machine learning models using TF-IDF without SMOTE.

Models	Accuracy	Precision	Recall	F1 score
RF	0.71	0.70	0.71	0.70
AB	0.64	0.65	0.64	0.64
ET	0.71	0.70	0.71	0.70
LR	0.73	0.73	0.73	0.72
MLP	0.71	0.71	0.71	0.71
GBM	0.70	0.70	0.70	0.70
kNN	0.66	0.65	0.66	0.65
ER-VC	0.72	0.72	0.72	0.71

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; TF-IDF: Term Frequency-Inverse Document Frequency; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier.

Table 6. Classification results of machine learning models using BoW without SMOTE.

Models	Accuracy	Precision	Recall	F1 score
RF	0.71	0.71	0.71	0.70
AB	0.68	0.69	0.68	0.68
ET	0.73	0.73	0.73	0.72
LR	0.72	0.72	0.72	0.71
MLP	0.71	0.71	0.71	0.71
GBM	0.73	0.73	0.73	0.72
kNN	0.52	0.55	0.52	0.50
ER-VC	0.74	0.74	0.74	0.74

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; BoW: Bag of Words; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier.

Table 7. Classification results of machine learning models using GloVe without SMOTE.

Models	Accuracy	Precision	Recall	F1 score
RF	0.60	0.59	0.59	0.59
AB	0.57	0.55	0.55	0.54
LR	0.59	0.58	0.59	0.55
MLP	0.65	0.63	0.65	0.63
ET	0.61	0.59	0.59	0.58
GBM	0.57	0.57	0.57	0.57
kNN	0.55	0.54	0.55	0.54
ER-VC	0.60	0.59	0.60	0.57

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; GloVe: Global Vectors; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier.

Table 8. Classification results of machine learning models using TF-IDF with SMOTE.

Models	Accuracy	Precision	Recall	F1 score
RF	0.81	0.82	0.81	0.81
AB	0.71	0.72	0.71	0.71
ET	0.82	0.83	0.82	0.82
LR	0.82	0.82	0.82	0.82
MLP	0.81	0.81	0.81	0.81
GBM	0.80	0.81	0.80	0.80
kNN	0.64	0.73	0.64	0.55
ER-VC	0.85	0.85	0.85	0.84

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; TF-IDF: Term Frequency-Inverse Document Frequency; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier.

models are trained to utilize three feature representation methods on an imbalanced and SMOTE-balanced dataset. Feature representation methods, including TF-IDF, BoW, and GloVe have been chosen as they show remarkable results in text classification. Accordingly, we selected the

Table 9. Classification results of machine learning models using BoW with SMOTE.

Models	Accuracy	Precision	Recall	F1 score
RF	0.78	0.79	0.78	0.78
AB	0.73	0.75	0.73	0.74
ET	0.78	0.78	0.78	0.78
LR	0.79	0.79	0.79	0.79
MLP	0.75	0.75	0.75	0.75
GBM	0.77	0.78	0.77	0.77
kNN	0.60	0.70	0.60	0.55
ER-VC	0.81	0.81	0.81	0.81

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; BoW: Bag of Words; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier.

most relevant ML models to classify symptoms. Machine learning models include RF, LR, MLP, GBM, AB, kNN, ET, and ER-VC. Experiments are performed to identify the most effective combination of feature extraction methods with ML models to classify symptoms into 'recovered', 'not recovered', or 'not survived'.

Results for scenario 1

At first, experiments have been performed on an imbalanced dataset using TF-IDF, BoW, and GloVe. Results of the proposed voting classifier are compared with the other baseline classifiers in terms of multiclass classification. Results presented in Table 5 show that LR achieves the highest results with a 0.73 accuracy score using TF-IDF on the imbalanced dataset. However, ER-VC achieved a 0.72 accuracy score, which is the second-highest among all classifiers. It can be noticed that RF, ET, and MLP achieve a 0.71 accuracy value. Moreover, AB shows the worst result with a 0.64 accuracy value using TF-IDF on the imbalanced dataset. AB often cannot generalize well in the case of an imbalanced dataset.

Results presented in Table 6 indicate that using BoW as a feature representation method improves the results of most of the classifiers on the imbalanced dataset. From Table 6, it can be observed that BoW does not improve the performance of MLP and kNN. The proposed voting classifier, ER-VC achieves a 0.74 accuracy score using BoW which is 2% higher than what is achieved by TF-IDF using an imbalanced dataset.

Table 7 shows the results of ML models when combined with GloVe features for the classification of an imbalanced dataset. A significant drop in the performance of ML classifiers can be observed. However, MLP yields the highest accuracy score of 0.65 whereas, the proposed ER-VC model does not perform well and acquired a 0.60 accuracy with GloVe features.

Results for scenario 2

The second scenario deals with the problem of imbalanced class distribution by the implementation of SMOTE. Data instances of the minority class are increased by oversampling to make a balanced dataset. Afterwards, ML models have been trained using TF-IDF, BoW, and GloVe on SMOTE-balanced datasets. The results of ML models using TF-IDF are presented in Table 8. It can be seen that SMOTE significantly improves the performance of ML models. As revealed by the results, SMOTE contributes to improving the models' classification results, and six out of eight models achieved higher than 80% results. SMOTE increases data instances of minority class by considering their distance to the k nearest neighbours of the minority class. In this way, the size of the minority class is increased by adding new data samples and making them appropriate for the training of the models. Hence the proposed voting classifier, ER-VC, which combines LR and ET outperforms other models and carries out prediction tasks with 0.85 accuracy, 0.85 precision, 0.85 recall, and 0.84 F1 scores.

Table 10. Classification results of machine learning models using GloVe with SMOTE.

Models	Accuracy	Precision	Recall	F1 score
RF	0.73	0.73	0.73	0.73
AB	0.58	0.58	0.58	0.58
ET	0.75	0.75	0.75	0.75
LR	0.60	0.59	0.60	0.59
MLP	0.65	0.67	0.65	0.69
GBM	0.63	0.63	0.63	0.63
kNN	0.64	0.64	0.64	0.63
ER-VC	0.73	0.73	0.73	0.73

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; GloVe: Global Vectors; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier.

Furthermore, the ML models are trained on the BoW feature representation technique. The performance of the models is compared in terms of classification results. Results shown in Table 9 prove that ML models using BoW do not achieve as robust results as achieved using TF-IDF on the SMOTE-balanced dataset.

Finally, ML models are combined with GloVe features for the classification of adverse reactions. The results reveal an overall decrease in the performance of ML models as shown in Table 10. However, a significant improvement in the results can be observed on the SMOTE-balanced dataset as compared to the performance of ML models when integrated with GloVe features on imbalanced data. Consequently, it proves that the BoW and GloVe feature representation techniques are not very effective in improving the performance of the models on the SMOTE-balanced dataset. However, SMOTE significantly improves the performance of ML models in classifying adverse events as 'not-survived', 'recovered', and 'not recovered'.

Performance analysis of ML models using different features

Figure 6(a) presents the accuracy comparison of ML models using BoW, TF-IDF, and GloVe without SMOTE while Figure 6(b) shows the performance comparison of ML models using BoW, TF-IDF, and GloVe using the SMOTE-balanced data. It can be observed that a substantial improvement in the accuracy of ML models occurred when they are trained using the SMOTE data.

Figure 7(a) presents the accuracy comparison of ML models using TF-IDF with and without SMOTE, Figure 7(b) presents the accuracy comparison of ML models using BoW with and without SMOTE while Figure 7(c) shows the accuracy comparison of ML models using GloVe with and without SMOTE. It shows that the results obtained by using BoW on the SMOTE-balanced dataset are better than the results achieved by using BoW on the imbalanced dataset. On the other side, the results of the models using BoW on the SMOTE-balanced dataset are 4% lower than the results obtained by using TF-IDF on the SMOTE-balanced dataset.

Performance comparison with deep neural networks

To substantiate the performance of the proposed voting classifier, it is also compared with deep learning models. We have used three deep learning models for experiments including LSTM⁵³, CNN⁵⁴, CNN-LSTM⁵⁶, and BiLSTM⁵⁵ for comparison purposes. Layered architecture and hyperparameter values are presented in Figure 8. The architecture of these models is based on the best results and optimized hyperparameters.

The same training and test split ratios are used for deep learning models. The deep learning models are used for

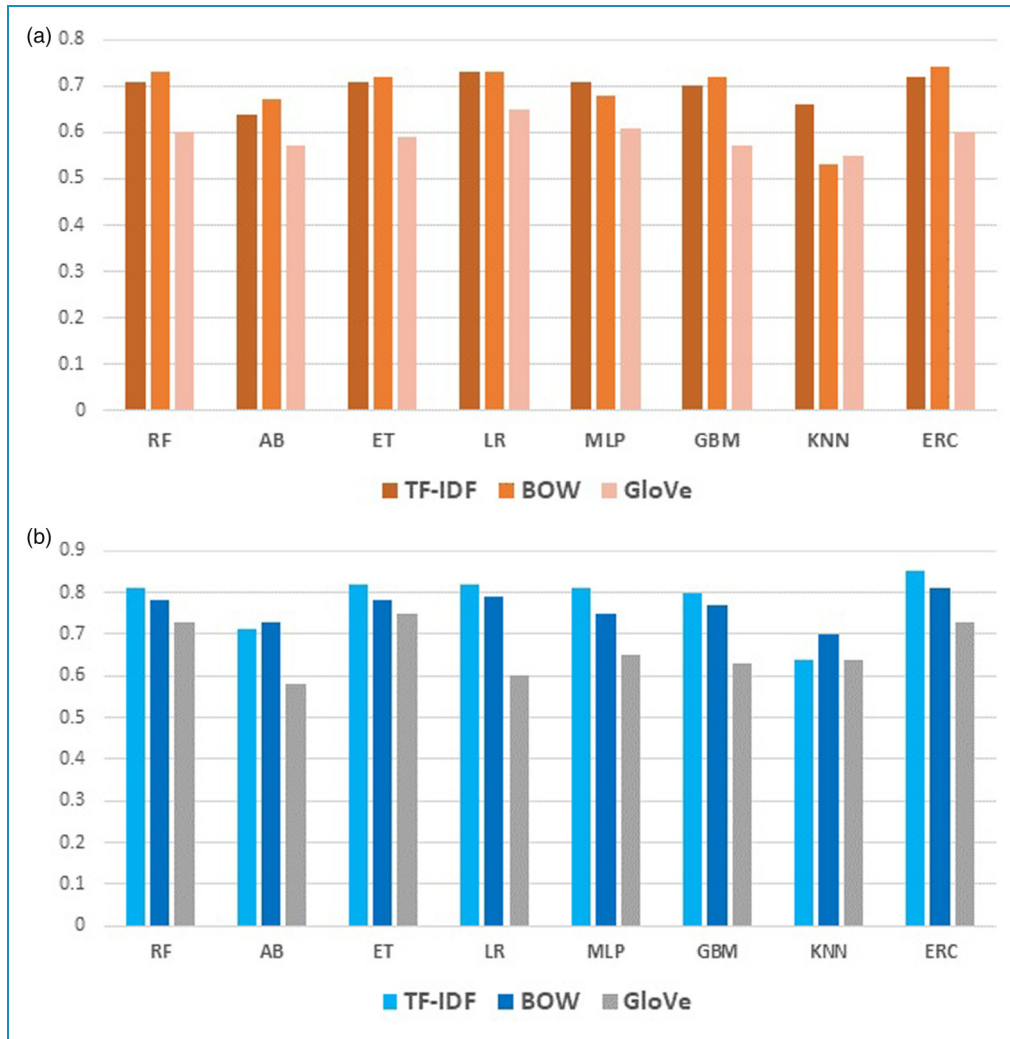


Figure 6. Performance analysis of ML models, (a) accuracy using TF-IDF, BoW, and GloVe without SMOTE, and (b) accuracy using TF-IDF, BoW, and GloVe using SMOTE. TF-IDF: Term Frequency-Inverse Document Frequency; BoW: Bag of Words; GloVe: Global Vectors; SMOTE: Synthetic Minority Oversampling Approach.

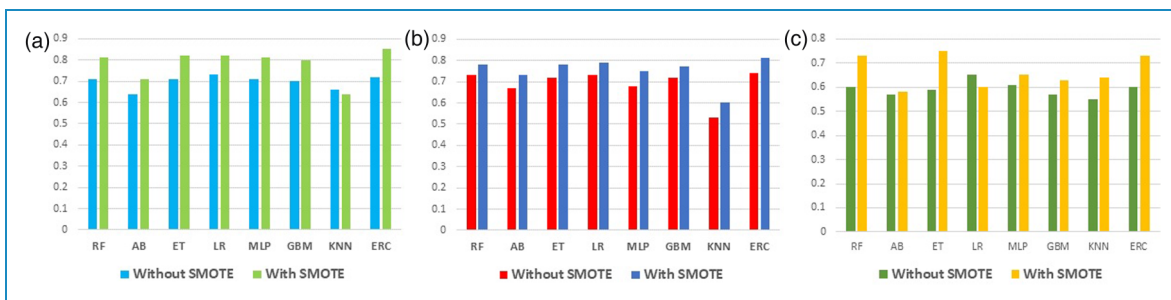


Figure 7. Performance analysis of ML models, (a) accuracy using TF-IDF with and without SMOTE, (b) accuracy using BoW with and without SMOTE, and (c) accuracy using GloVe with and without SMOTE. TF-IDF: Term Frequency-Inverse Document Frequency; BoW: Bag of Words; GloVe: Global Vectors; SMOTE: Synthetic Minority Oversampling Approach.

experiments considering both the original and the SMOTE-balanced datasets. The training and testing accuracy curve of the used deep learning models is shown in Figure 9.

Classification results of deep learning models with and without SMOTE are presented in Table 11. It can be observed that LSTM achieves the highest result with a

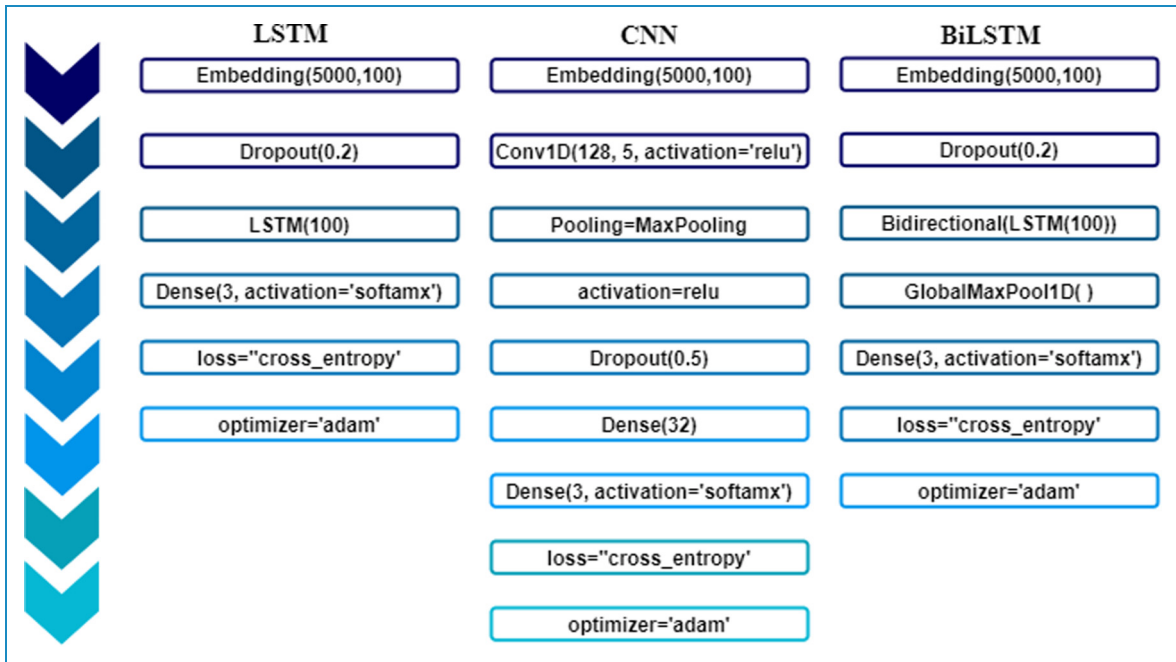


Figure 8. Layered architecture of the deep neural networks.

0.70 value of accuracy, precision, recall, and F1 score without using SMOTE. CNN has shown the lowest result on the imbalanced dataset. Given the small size of training data available for the deep neural networks, the performance is not good. However, using the SMOTE-balanced dataset, CNN-LSTM has achieved the highest accuracy score of 0.82 followed by similar precision, recall, and F1 score. LSTM and CNN have yielded 0.81 accuracy, precision, recall, and F1 scores. However, these values are lower than the proposed model, namely ER-VC. Despite that, results for deep learning models confirm that SMOTE has significantly improved the performance of CNN-LSTM, LSTM, and CNN models while BiLSTM has achieved similar results with and without SMOTE.

Results with data splitting prior to SMOTE

To show the significance of the proposed model, this study also deployed another approach where SMOTE technique is used with a training set only. The data is split into training and test subsets and SMOTE is applied only to the training set to balance the samples of different classes. The results of machine learning models given in Table 12 reveal a drop in the performance of learning models; however, the proposed model ER-VC still shows better results with this approach. ER-VC outperforms other models with a 0.75 accuracy score in comparison with all other used models.

Table 13 presents the results of deep learning models when trained with SMOTE-balanced data and tested with original data. A notable decline in the performance of

models is discerned. However, in this case, as well, the performance of deep learning models did not exceed the performance of our proposed ER-VC classifier.

Validation of proposed approach for binary classification

The current study validates the proposed ER-VC model by predicting the survival status of the vaccinated individuals. In accordance with this, we integrated 'SYMPTOM_TEXT' as features and 'DIED' as the target class. It involves a total of 5351 data instances among which 810 are labeled as Y (not survived) and the remainder of the records are labeled as N (survived). The proposed ER-VC model is trained on 80% train data which is pre-processed and balanced using SMOTE. Experimental results after testing ER-VC on binary classification are shown in Table 14. Empirical results showed that the proposed ER-VC model manifested state-of-the-art performance in the prognosis of death risks by analyzing the adverse events reported to VAERS. Concerning the feature set, TF-IDF leads with a 0.98 accuracy score with its ability to extract features with more predictive information regarding target variables as compared to BoW which only provides a feature set of terms irrespective of their importance in the document, and GloVe which is inefficient when it comes to unknown words.

Figure 10 demonstrates the number of instances predicted correctly following the given target variable. It can be observed that ER-VC wrongly predicted only 30

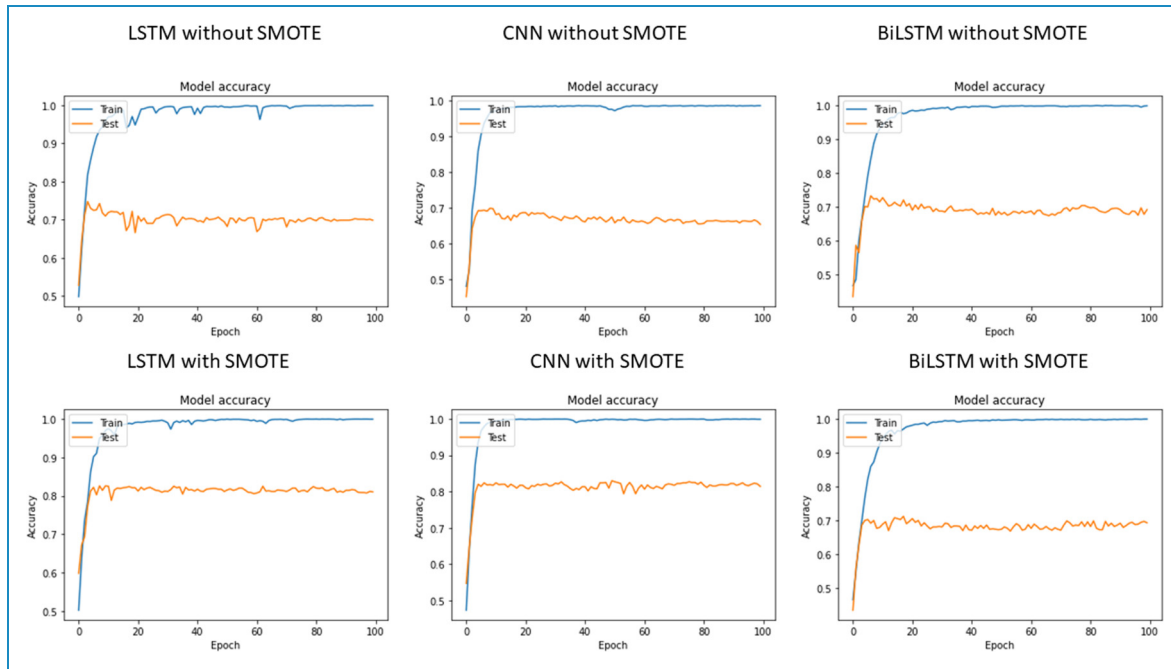


Figure 9. Accuracy measure of deep neural networks with respect to each epoch.

Table 11. Classification results of deep neural networks without SMOTE.

Dataset	Models	Acc.	Prec.	Rec.	F1
No SMOTE	LSTM	0.70	0.70	0.70	0.70
	CNN	0.64	0.65	0.65	0.64
	CNN-LSTM	0.67	0.67	0.67	0.67
	BiLSTM	0.69	0.69	0.69	0.69
SMOTE	LSTM	0.81	0.81	0.81	0.81
	CNN	0.81	0.81	0.81	0.81
	CNN-LSTM	0.82	0.82	0.82	0.82
	BiLSTM	0.69	0.69	0.69	0.69

LSTM: Long Short Term Memory; CNN: Convolutional Neural Network; BiLSTM: Bidirectional LSTM; SMOTE: Synthetic Minority Oversampling Approach.

instances from a total of 1817 instances when integrated with TF-IDF features as shown in Figure 10(a). Contrarily, Figure 10(b) shows that ER-VC in combination with BoW features made 64 wrong predictions out of 1817 instances. Whereas, in the case of GloVe features, the wrong predictions totals 162 which shows its poor performance in binary classification as presented in Figure 10(c). BoW generates features irrespective of their importance

concerning the target class whereas TF-IDF with its ability to extract features that are significant relative to the analysis excels in its performance. This resulted in an effective and robust prognosis of death risks following the COVID-19 vaccine using the proposed ER-VC model combined with TF-IDF features.

To further show the significance of the validation, we also conducted experiments by applying SMOTE on the training set only for the binary classification. Table 15 shows that the performance of the model follows a similar trend as shown in Table 14.

Conclusion

The COVID-19 vaccine has caused different symptoms and adverse reactions in different individuals, ranging from mild to severe, and many deaths have also been reported post-COVID-19 vaccination. Analyzing the post-vaccination symptom can play an important role to understand the relation between different symptoms and fatality, thereby helping the health professionals escalate serious patients and take timely precautionary measures. This study proposes a framework to analyze the adverse events caused by the COVID-19 vaccine leading to death so that health professionals are alerted beforehand. The proposed model predicted three significant events including 'not survived', 'recovered', and 'not recovered' based on the adverse events followed by the second dosage of the COVID-19 vaccine. Keeping in view the data imbalance, experiments are performed using the original dataset, as

Table 12. Results of machine learning models with data split before applying SMOTE.

Model	Accuracy	Precision	Recall	F1 score
TF-IDF				
RF	0.71	0.71	0.71	0.71
AB	0.67	0.69	0.67	0.68
ET	0.72	0.72	0.72	0.72
LR	0.74	0.74	0.74	0.74
MLP	0.70	0.69	0.70	0.70
GBM	0.71	0.72	0.71	0.71
KNN	0.52	0.70	0.52	0.42
ER-VC	0.75	0.75	0.75	0.75
GloVe				
RF	0.61	0.61	0.61	0.61
AB	0.54	0.55	0.54	0.54
ET	0.58	0.57	0.58	0.58
LR	0.60	0.60	0.60	0.59
MLP	0.62	0.62	0.62	0.62
GBM	0.54	0.54	0.54	0.54
KNN	0.55	0.57	0.55	0.55
ER-VC	0.63	0.62	0.63	0.62
BoW				
RF	0.69	0.71	0.69	0.70
AB	0.67	0.69	0.67	0.68
ET	0.58	0.57	0.58	0.58
LR	0.73	0.73	0.73	0.73
MLP	0.62	0.62	0.62	0.62
GBM	0.72	0.73	0.72	0.72
KNN	0.55	0.57	0.55	0.55
ER-VC	0.73	0.74	0.73	0.74

RF: Random Forest; LR: Logistic Regression; MLP: Multilayer Perceptron; GBM: Gradient Boosting Machine; AB: AdaBoost, kNN: k Nearest Neighbours; ET: Extra Tree Classifier; GloVe: Global Vectors; SMOTE: Synthetic Minority Oversampling Approach; ER-VC: Extreme Regression-Voting Classifier; TF-IDF: Term Frequency-Inverse Document Frequency; BoW: Bag of Words.

Table 13. Deep learning models' results with data splitting prior to SMOTE.

Model	Accuracy	Precision	Recall	F1 score
LSTM	0.61	0.62	0.61	0.61
CNN	0.61	0.61	0.61	0.61
CNN-LSTM	0.63	0.63	0.63	0.63
BiLSTM	0.61	0.61	0.61	0.61

LSTM: Long Short Term Memory; CNN: Convolutional Neural Network; BiLSTM: Bidirectional LSTM; SMOTE: Synthetic Minority Oversampling Approach.

Table 14. Classification results of proposed ER-VC model for binary classification.

Feature	Acc.	Class	Prec.	Rec.	F1
BoW	0.96	Survived	0.97	0.96	0.97
		Not-survived	0.96	0.97	0.96
		Weighted avg	0.96	0.96	0.96
TF-IDF	0.98	Survived	0.98	0.98	0.98
		Not-survived	0.98	0.98	0.98
		Weighted avg	0.98	0.98	0.98
GloVe	0.91	Survived	0.93	0.89	0.91
		Not-survived	0.89	0.93	0.91
		Weighted avg	0.91	0.91	0.91

GloVe: Global Vectors; ER-VC: Extreme Regression-Voting Classifier; TF-IDF: Term Frequency-Inverse Document Frequency; BoW: Bag of Words.

well as, the SMOTE-balanced dataset. The efficacy of the proposed voting classifier ER-VC is investigated in comparison with many well-known machine learning using TF-IDF, BoW, and GloVe, and deep learning models. After extensive experiments, it is concluded that BoW and GloVe are not effective for the classification of COVID-19 vaccine symptoms. TF-IDF, on the other hand, has shown significant improvement in the classification of vaccine symptoms when it is applied to the SMOTE-balanced dataset. Experimental results proved that the proposed voting classifier surpassed other models with a 0.85 accuracy score using TF-IDF on the SMOTE-balanced dataset. Moreover, the comparison concerning the benchmark state-of-the-art deep neural

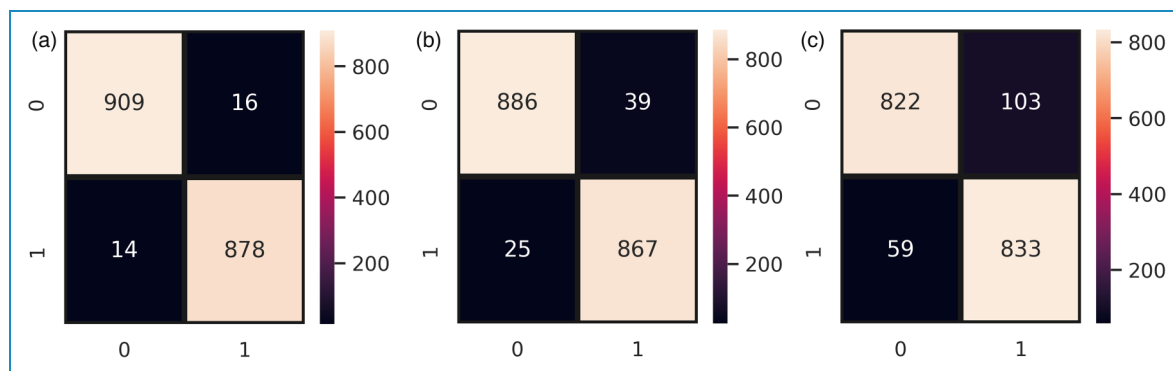


Figure 10. Confusion matrix of ER-VC concerning binary classification, (a) ER-VC with TF-IDF, (b) ER-VC with BoW, and (c) RT-VC with GloVe. GloVe: Global Vectors; ER-VC: Extreme Regression-Voting Classifier; TF-IDF: Term Frequency-Inverse Document Frequency; BoW: Bag of Words.

Table 15. Classification results of proposed ER-VC model for binary classification with SMOTE-balanced train set.

Feature	Acc.	Class	Prec.	Rec.	F1
BoW	0.94	Survived	0.97	0.96	0.96
		Not-survived	0.79	0.83	0.81
		Weighted avg	0.94	0.94	0.94
TF-IDF	0.96	Survived	0.97	0.98	0.98
		Not-survived	0.90	0.85	0.88
		Weighted avg	0.96	0.96	0.96
GloVe	0.86	Survived	0.95	0.88	0.92
		Not-survived	0.53	0.74	0.62
		Weighted avg	0.89	0.86	0.87

GloVe: Global Vectors; ER-VC: Extreme Regression-Voting Classifier; TF-IDF: Term Frequency-Inverse Document Frequency; BoW: Bag of Words; SMOTE: Synthetic Minority Oversampling Approach.

networks confirms the performance of ER-VC better is significantly better than deep learning models. Furthermore, the effectiveness of the proposed model has been proved by experiments for binary classification where the model shows robust results with a 0.98 accuracy score. Machine learning models and deep neural networks tend to perform better given a larger dataset, therefore, in the future, we plan to incorporate a larger dataset for more accurate results.

Contributorship: Eysha Saad (ES) and Saima Sadiq (SS) conceived the idea and performed analysis. Ramish Jamil (RJ) and Furqan Rustam (FR) performed the data curation and formal

analysis. Arif Mehmood (AM) and Gyu Sang Choi (GSC) provided the resources and software. Imran Ashraf (IA) supervised the work. SS, RJ, and FR conducted experiments. ES wrote the initial manuscript, IA did the write-review & editing. All authors reviewed the manuscript.

Declaration of Conflicting Interests: The authors declare that there is no conflict of interest.

Ethical approval: Not Applicable.

Funding: This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2019R1A2C1006159 and Grant NRF-2021R1A6A1A03039493.

Guarantor: Not applicable.

ORCID iD: Imran Ashraf  <https://orcid.org/0000-0002-8271-6496>

References

1. WHO. Coronavirus disease (Covid-19), 2021. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
2. Lv H, Wu NC and Mok CK. Covid-19 vaccines: Knowing the unknown. *Eur J Immunol* 2020; 507: 939–943.
3. WHO. Estimating mortality from covid-19. *Scientific brief, 4 August 2020. Technical Report, World Health Organization, 2020*.
4. Shah A, Marks PW and Hahn SM. Unwavering regulatory safeguards for covid-19 vaccines. *Jama* 2020; 32410: 931–932.
5. Voysey M, Clemens SAC, Madhi SA, et al. Safety and efficacy of the chadox1 ncov-19 vaccine (azd1222) against sars-cov-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* 2021; 39710269: 99–111.

6. Corey L, Mascola JR, Fauci AS, et al. A strategic approach to covid-19 vaccine r&d. *Science* 2020; 3686494: 948–950.
7. Hussain S, Muhammad L, Ishaq F, et al. Performance evaluation of various data mining algorithms on road traffic accident dataset. In *Information and Communication Technology for Intelligent Systems*. Springer, 2019. pp. 67–78.
8. Jagadeesh K and Rajendran A. Machine learning approaches for analysis in healthcare informatics 2021; pp. 105–122.
9. Ruan Q, Yang K, Wang W et al. Clinical predictors of mortality due to Covid-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* 2020; 465: 846–848.
10. Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020; 1807: 934–943.
11. Charumilind S, Craven M, Lamb J, et al. When will the Covid-19 pandemic end? an update, 2021.
12. Costanzo M, De Giglio MAR and Roviello GN. Anti-coronavirus vaccines: Past investigations on sars-cov-1 and mers-cov, the approved vaccines from biontech/pfizer, moderna, oxford/astrazeneca and others under development against sarscov-2 infection. *Curr Med Chem* 2021.; 29: 4–18
13. Matrajt L, Eaton J, Leung T, et al. Vaccine optimization for covid-19: Who to vaccinate first? *Sci Adv* 2020; 76: eabf1374.
14. Umer M, Ashraf I, Ullah S et al. Covinet: a convolutional neural network approach for predicting Covid-19 from chest x-ray images. *J Ambient Intell Humaniz Comput* 2021; 13: 535–547.
15. Ashraf I, Alnumay WS, Ali R, et al. Prediction models for Covid-19 integrating age groups, gender, and underlying conditions. *Comput, Materials Continua* 2021; 673: 3009–3044.
16. Cheng Y, Luo R, Wang K, et al. Kidney disease is associated with in-hospital death of patients with Covid-19. *Kidney Int* 2020; 975: 829–838.
17. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with Covid-19 in wuhan, china: a retrospective cohort study. *Lancet* 2020; 39510229: 1054–1062.
18. Jiang M, Li C, Zheng L, et al. A biomarker-based age, biomarkers, clinical history, sex (abcs)-mortality risk score for patients with coronavirus disease 2019. *Ann Transl Med* 2021; 93: 1–5.
19. Onan A. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *J King Saud Univ-Comput Inform Sci* 2022; 34: 2098–2117.
20. Onan A and Toçoğlu MA. A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access* 2021; 9: 7701–7722.
21. Onan A, Korukoğlu S and Bulut H. Ensemble of keyword extraction methods and classifiers in text classification. *Expert Syst Appl* 2016; 37: 232–247.
22. Onan A. An ensemble scheme based on language function analysis and feature engineering for text genre classification. *J Inf Sci* 2018; 44: 28–47.
23. Onan A. A feature selection model based on genetic rank aggregation for text sentiment classification. *J Inf Sci* 2017; 43: 25–38.
24. Onan A, Korukoğlu S and Bulut H. A hybrid ensemble pruning approach based on consensus clustering and multi-objective evolutionary algorithm for sentiment classification. *Inf Process Manag* 2017; 53: 814–833.
25. Onan A. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurr Comp: Pract Exp* 2021; 33: e5909.
26. Onan A. Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. *Comput Appl Eng Educ* 2021; 29: 572–589.
27. Onan A. Mining opinions from instructor evaluation reviews: a deep learning approach. *Comput Appl Eng Educ* 2020; 28: 117–138.
28. Onan A. Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering. *IEEE Access* 2019; 7: 145614.
29. Onan A. Biomedical text categorization based on ensemble pruning and optimized topic modelling. *Comput Math Methods Med* 2018; 2018: 2497471.
30. Onan A. Topic-enriched word embeddings for sarcasm identification. *Comput Sci On-line Conference* 2019; 984: 293–304.
31. Sadiq S, Mehmood A and Ullah S. Aggression detection through deep neural model on twitter. *Future Gener Comput Syst* 2021; 114: 120–129.
32. Sadiq S, Khalid MU, Ullah S, et al. Classification of β -thalassemia carriers from red blood cell indices using ensemble classifier. *IEEE Access* 2021; 9: 45528–45538.
33. Castiglione A, Vijayakumar P, Nappi M, et al. Covid-19: automatic detection of the novel coronavirus disease from ct images using an optimized convolutional neural network. *IEEE Trans Ind Inf* 2021; 17: 6480–6488.
34. Castiglione A, Umer M, Sadiq S, et al. The role of internet of things to control the outbreak of Covid-19 pandemic. *IEEE Int Things J* 2021; 8: 16072–16082.
35. Estiri H, Strasser ZH, Klann JG, et al. Predicting Covid-19 mortality with electronic medical records. *NPJ Digit Med* 2021; 41: 1–10.
36. Garg A. Covid-19 world vaccine adverse reactions, 2021. <https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions?select=2021VAERSDATA.csv>.
37. VAERS. COVID-19 World Vaccine Adverse Reactions. <https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions?select=2021VAERSDATA.csv>, 2021. [Online; accessed September 06, 2021].
38. Zhang P, Gao D, An K, et al. A programmable polymer library that enables the construction of stimuli-responsive nanocarriers containing logic gates. *Nat Chem* 2020; 124: 381–390.
39. Giveki D. Scale-space multi-view bag of words for scene categorization. *Multimed Tools Appl* 2021; 801: 1223–1245.
40. Meijer H, Truong J and Karimi R. Document embedding for scientific articles: Efficacy of word embeddings vs tfidf. *arXiv preprint arXiv:210705151* 2021.
41. Obayes HK, Al-Turaihi FS and Alhussayni KH. Sentiment classification of user's reviews on drugs based on global vectors for word representation and bidirectional long short-term memory recurrent neural network. *Indones J Elect Eng Comput Sci* 2021; 231: 345–353.
42. Batista GE, Prati RC and Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor Newsl* 2004; 61: 20–29.
43. Chawla NV, Bowyer KW, Hall LO, et al. Smote: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16: 321–357.

44. Ishaq A, Sadiq S, Umer M, et al. Improving the prediction of heart failure patients' survival using smote and effective data mining techniques. *IEEE Access* 2021; 9: 39707–39716.
 45. Biau G and Scornet E. A random forest guided tour. *Test* 2016; 252: 197–227.
 46. Akhter I, Jalal A and Kim K. Pose estimation and detection for event recognition using sense-aware features and adaboost classifier. In *2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST)*. IEEE, pp. 500–505.
 47. Sharaff A and Gupta H. Extra-tree classifier with metaheuristics approach for email classification. In *Advances in Computer Communication and Computational Sciences*. Springer, 2019. pp. 189–197.
 48. Saad E, Din S, Jamil R, et al. Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums. *IEEE Access* 2021; 9: 85721–85737.
 49. Gardner MW and Dorling S. Artificial neural networks (the multilayer perceptron): a review of applications in the atmospheric sciences. *Atmos Environ* 1998; 32:14-15: 2627–2636.
 50. Natekin A and Knoll A. Gradient boosting machines, a tutorial. *Front Neurorobot* 2013; 7: 21.
 51. Sharma KK and Seal A. Spectral embedded generalized mean based k-nearest neighbors clustering with s-distance. *Expert Syst Appl* 2021; 169: 114326.
 52. Landi F, Baraldi L, Cornia M, et al. Working memory connections for lstm. *Neural Netw* 2021; 144: 334–341.
 53. Jamil R, Ashraf I, Rustam F, et al. Detecting sarcasm in multi-domain datasets using convolutional neural networks and long short term memory network model. *PeerJ Comput Sci* 2021; 7: e645.
 54. Liu B, Song C, Wang Q, et al. Forecasting of china's solar pv industry installed capacity and analyzing of employment effect: based on gra-bilstm model. *Environ Sci Poll Res* 2021; 29: 1–17.
 55. Wang J, Yu L, Lai K, et al. Dimensional sentiment analysis using a regional CNN-LSTM model. Proceedings of the 54th annual meeting of the association for computational linguistics, August 7-12, Berlin, Germany; 2016; pp. 225–230.
-