

A textual-based featuring approach for depression detection using machine learning classifiers and social media texts

by Perpustakaan Referensi

Submission date: 29-Apr-2025 09:49AM (UTC+0700)

Submission ID: 2660375963

File name: CBM_1-s2.0-S0010482521002936-main.pdf (6.72M)

Word count: 8807

Character count: 46756



10

A textual-based featuring approach for depression detection using machine learning classifiers and social media texts

Raymond Chiong^{a,*}, Gregorius Satia Budhi^{a,b,**}, Sandeep Dhakal^a, Fabian Chiong^c

^a School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, 2308, Australia

^b Informatics Department, Petra Christian University, Surabaya, 60236, Indonesia

^c Alice Springs Hospital, The Gap, NT, 0870, Australia

ARTICLE INFO

Keywords:

Depression detection
Social media
Textual-based featuring
Machine learning
Imbalanced data

ABSTRACT

Depression is one of the leading causes of suicide worldwide. However, a large percentage of cases of depression go undiagnosed and, thus, untreated. Previous studies have found that messages posted by individuals with major depressive disorder on social media platforms can be analysed to predict if they are suffering, or likely to suffer, from depression. This study aims to determine whether machine learning could be effectively used to detect signs of depression in social media users by analysing their social media posts—especially when those messages do not explicitly contain specific keywords such as 'depression' or 'diagnosis'. To this end, we investigate several text preprocessing and textual-based featuring methods along with machine learning classifiers, including single and ensemble models, to propose a generalised approach for depression detection using social media texts. We first use two public, labelled Twitter datasets to train and test the machine learning models, and then another three non-Twitter depression-class-only datasets (sourced from Facebook, Reddit, and an electronic diary) to test the performance of our trained models against other social media sources. Experimental results indicate that the proposed approach is able to effectively detect depression via social media texts even when the training datasets do not contain specific keywords (such as 'depression' and 'diagnose'), as well as when unrelated datasets are used for testing.

1. Introduction

According to the World Health Organisation, depression is the most prevalent mental disorder, which affects more than 300 million people worldwide [1]. Depression is also the leading cause of more than two-thirds of suicides every year [2]. However, due to self-denial among some patients and poor recognition of the issue in many places, depression can remain undiagnosed or untreated. The lack of diagnosis and treatment can further aggravate the condition [3], which could lead to reduced quality of life and, in acute cases, an inability to maintain employment [4,5].

Numerous studies in the literature agree that social media platforms, where people freely share their thoughts and express their feelings, could be a vital source for monitoring health issues and trends [6,7]. Posts on platforms, such as Twitter and Facebook, enable researchers to investigate multiple aspects of psychological concerns and human

behaviour [8,9]. Studies focusing on mental depression, for example, have found that tweets posted by individuals with a major depressive disorder could be utilised to predict future episodes of depression in those individuals [3]. A recent survey indicated that an increasing number of people with depression symptoms, especially teenagers and young adults, are turning to social media to express their feelings (see Fig. 1) [10]. Related work in this domain, however, often relies on specific keywords such as 'depression' and 'diagnose' when utilising the data, while the fact is that social media users suffering from depression are not likely to use such words directly.

Given the prevalence of depression, its impacts, and the potential of using social media texts for predicting depression, the main goal of this study is to analyse whether machine learning (ML) methods could be effectively used to detect depression in people by analysing their social media texts, but without relying on specific keywords. Social media texts are often unstructured, and therefore ML, which is good at dealing with

* Corresponding author.

** Corresponding author. School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, 2308, Australia.

E-mail addresses: raymond.chiong@newcastle.edu.au (R. Chiong), gregorius.satiabudhi@uon.edu.au, greg@petra.ac.id (G.S. Budhi), sandeep.dhakal@newcastle.edu.au (S. Dhakal), fabian.chiong@nt.gov.au (F. Chiong).

<https://doi.org/10.1016/j.combiomed.2021.104499>

Received 12 February 2021; Received in revised form 12 May 2021; Accepted 14 May 2021

Available online 17 May 2021

0010-4825/© 2021 Elsevier Ltd. All rights reserved.

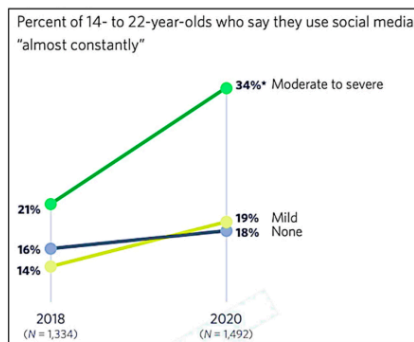


Fig. 1. Frequency of social media use, by depressive symptom levels, 2018–2020 [10].

nonlinearity, is expected to be a better option than traditional statistical methods for analysing them. Even though we train the ML models using Twitter posts in this study, they should be equally applicable to other text-based messages. Specifically, the models are first trained and tested using two publicly labelled datasets comprising of Twitter tweets [11, 12]—with and without specific keywords such as ‘depression’ and ‘diagnose’. The models are then re-tested using additional non-Twitter depression-class-only datasets [13–15], primarily to investigate the performance of our trained models in detecting depression in people who posted messages on various social media platforms. More precisely, the performance of our generalised approach is investigated against five different, publicly available datasets comprising of social-media text from various sources, including Twitter, Facebook, Reddit, and a personal electronic diary. This approach contrasts with most approaches in the literature that use parts of the same datasets for both training and testing the ML models, as well as rely on specific keywords.

Our detection model is constructed using ML methods combined with textual-based featurizing; textual-based featurizing extracts input features from the text itself and is, thus, more independent of the system than other featurizing processes [16]. We apply ML classifiers—both single and ensemble models—that are widely used in solving prediction problems; these methods have been selected because of their excellent prediction performance in previous studies [16,17]. The performance of these models is, however, dependent on the type of data and features used for training them. Therefore, we investigate the performance of our models using datasets from several social media texts to verify that the proposed approach is a general one. Additionally, the features used for training are preprocessed and extracted using a combination of various methods that performed well in previous work [16–18]. More specifically, text preprocessing methods, such as tokenisation, stop word removal, detection of negation words, correction of elongation words, and part of speech (POS) lemmatisation, are applied. Regarding the input features, we implement a bag-of-words (BOW) feature extraction method combined with count vectorisation and n-gram words (from unigram to trigram) to extract features. This work thus contributes to the relevant research areas of ML and natural language processing as well as the study of mental health problems by proposing a generalised approach for depression detection that is effective especially when the social media users are not aware of their depression or are in denial.

The rest of this paper is organised as follows. In Section 2, we review related work on depression detection. Then in Section 3, we describe, in detail, the design of the textual-based depression detection framework developed for this study. Experimental results and discussions are

presented in Section 4, and Section 5 concludes the paper and highlights potential future research directions.

2. Related work

Numerous studies on automatic detection of the symptoms of depression have been carried out using artificial intelligence methods such as ML. A major stream of research involves depression detection using medical data, such as fMRI signature [19], results of depression questionnaires (such as DASS21 and DASS42) [1,20], or clinical criteria for depression as defined in DSM-5 and ICD-10 [21]. Data from clinical interviews, using systems such as the Distress Analysis Interview Corpus-Wizard of Oz (DAIC-WOZ [22]) [2,23], has also been collected. DAIC-WOZ data includes videos, speeches, and text transcriptions of the participants, who could be either distressed or non-distressed. In recent years, researchers have also focused on depression detection using text messages from social media platforms, such as Twitter, Facebook, Reddit, and WeChat [3,6,9,11,24–27]—in the hope that social media texts can help detect depression even when the individual is unaware of their depression or is in denial.

The majority of studies on depression detection using social media messages usually follow either a textual-based featurizing approach or a person descriptive-based featurizing approach. Textual-based featurizing focuses on the linguistic features of the social media text, such as words, POS, n-gram, and other linguistic characteristics [3,6,9,25–28]. In contrast, the descriptive-based featurizing approach focuses on descriptions of the subject, such as age, gender, employment status, income, consumption of drugs or alcohol, smoking, and other details of the subject or patient [11,21,29–32]. These features are then input into the detection models. Most models for depression detection have been developed using ML classifiers, such as the Support Vector Machine (SVM), Multilayer Perceptron (MLP), Logistic Regression (LR), Decision Tree (DT), Naive Bayes (NB), Maximum Entropy (ME), K-Nearest Neighbours (KNN), Adaptive Boosting (AB), Random Forest (RF), Gradient Boosting (GB), Bagging Predictors (BP), and other single and ensemble models [1,3,6,19,21,26,29,30]. Deep learning methods, such as the Long Short-Term Memory (LSTM) [24] and Convolutional Neural Network (CNN) [9,33], have also been used. Additionally, several studies have constructed custom detectors [11,26,31]. A concise overview of related work from the past five years (2017–2021) is provided in Table 1.

In this paper, we discuss our efforts towards building a depression detection framework using textual features from social media posts. This is achieved by utilising some proven methods of preprocessing, featurizing and ML classifiers from our previous studies [16,17]. In research on depression detection using social media texts, custom datasets are frequently created but not made publicly available. In contrast, in this study, we utilise five publicly available datasets. These include two binary-class Twitter datasets that are used for both training and testing, and three single-class datasets from Facebook, Reddit and an electronic diary for further testing. While our goal is a generalised approach for depression detection in social media texts, the datasets include only text messages and exclude any emoticons, emojis, pictures, videos and web links that are commonly part of social media messages. Additionally, we address the issue of overfitting that generally arises when collecting depression data from social media messages. A model might perform poorly on datasets it was not trained on due to overfitting. We also focus our efforts on overcoming the problem of imbalanced data samples, which can negatively impact the performance of classification models.

3. Methods

We designed a framework for detecting depression through an analysis of only social media texts, by using methods inspired by previous studies in the literature. For the preprocessing of social media texts (see Fig. 2), we implemented several methods that performed well in

Table 1
An overview of related work from 2017 to 2021.

Authors	Year	Feature Type	Dataset sources	Methods ^a	Results ^b
Depression detection on social media messages					
Shen et al. [11]	2017	Descriptive	Twitter (Public/Pu [11])	MDL ^c , MSNL ^c , WDL ^c , NB	Acc: 85%; Pre: 85%; Rec: 85%; F1: 85%
Hassan et al. [6]	2017	Textual	Twitter (Private/Pr)	SVM, NB, ME	Acc: 91%; Pre: 83%; Rec: 79%
Chen et al. [24]	2018	Textual	Survey and WeChat (Pr)	LSTM	Present the results in several graphs
Islam et al. [25]	2018	Textual	Facebook (Pr)	DT, KNN, SVM, Ensemble	Pre: 59%; Rec: 97%; F1: 73%
Burdisso et al. [26]	2019	Textual	Reddit (Pr)	SS3 ^c , KNN, LR, SVM, NB	Pre: 63%; Rec: 60%; F1: 61%
Fatima et al. [27]	2019	Textual	Reddit (Pr)	MLP, SVM, LR	Acc: 91.63%; Pre: 91.83%; Rec: 91.85%
Lin et al. [33]	2020	Visual and Textual	Twitter (Pu [11]) and Images	CNN	Acc: 88.4%; Pre: 90.3%; Rec: 87%; F1: 93.6%
Alsagri and Mourad [3]	2020	Textual	Twitter (Pr)	SVM, NB, DT	Acc: 82.5%; Pre: 73.91%; Rec: 85%; F1: 79.06%; AUC: 0.78
Kim et al. [9]	2020	Textual	Reddit (Pu [9])	CNN, XGBoost	Acc: 75.13%; Pre: 89.1%; Rec: 71.75%; F1: 79.49%
Depression detection on other sources (non-social media)					
Jung et al. [31]	2017	Textual/Ontology	35 FAQs about depression from multiple sources	DT, LR	Acc: 75%; Pre: 76.1%
Samareh et al. [23]	2018	Audio, video and textual	DAIC-WOZ ^d (Pu [22])	RF	RMSE: 5.12; MAE: 4.12
Priya et al. [1]	2020	Descriptive	DASS-21 ^e (Pr)	NB, DT, RF, SVM, KNN	Acc: 85.5%; Pre: 82.2%; Rec: 85%; F1: 83.6
Kumar et al. [20]	2020	Descriptive	DASS-42 (Pr)	RBFN ^f , NB, KNN, MLP, RF, K-Star, J48	Acc: 96%; Pre: 96%; Rec: 96%; F1: 96%; AUC: 0.99
Srimadhur and Lalitha [2]	2020	Spectrogram & End-to-end	DAIC-WOZ (Pu [22])	CNN (Non-depressed class)	Pre: 65%; Rec: 92%; F1: 76%
Jothi et al. [30]	2020	Descriptive	Online survey (Pr)	J48, NB, RF	Acc: 95.7%; Rec: 97.5%; Spe: 86.3%
Filho et al. [29]	2021	Descriptive	Patients' clinical evaluation (Pr)	RF, LR, KNN, DT, AB, SVM, GB	Acc: 89%

^a The first method in bold is the method with the best result.

^b Acc = Accuracy; Pre = Precision; Rec = Recall; Spe = Specificity; F1 = F-measure; AUC = Area Under the Curve; RMSE = Root Mean Square Error; MAE = Mean Absolute Error.

^c SS3 = Sequential S3 (Smoothness, Significance, and Sanction); MDL = Multimodal Depressive Dictionary Learning; WDL = Wasserstein Dictionary Learning; MSNL = Multiple Social Networking Learning; fMRI = functional Magnetic Resonance Imaging; MLDA = Maximum Entropy Linear Discriminant Analysis; DASS = Depression, Anxiety and Stress Scale questionnaire; RBFN = Radial Basis Function Network; DSM-5 = Diagnostic and Statistical Manual of Mental Disorders 5th edition; ICD-10 = International Statistical Classification of Diseases and Related Health Problems nu. 10; BCC = Bayesian Classifier Chains; PCC = Probabilistic Classifier Chains; SCC = Super Class Classifier, Bagging, ECC = Ensemble of Classifier Chains; Pruned sets, CC = Classifier Chains; DAIC-WOZ = The Distress Analysis Interview Corpus Wizard of Oz.

previous work [16], as follows:

- Removal of punctuation, numbers, and stopwords.
- Word correction, which is an essential step in preprocessing, for reducing the diversity of features. The following three methods of word correction are utilised in this study:
 - Spelling error correction.** Detectors differentiate between misspelled words and their correct forms, thereby unnecessarily increasing the detection complexity. To avoid this, we detected and corrected misspelled words by utilising Peter Norvig's code for spelling correction [34] based on probability theory. In this method, the investigated words are compared against a large database of words, and the most probable replacements are chosen for misspelled words.
 - Elongated word correction.** Word diversity, and thereby the detection complexity, are also impacted by the presence of elongated words or word-stretchers such as 'yesss', 'fiifine' and 'youuu'. These were also reduced to their original forms using Peter Norvig's code for spelling correction [34].
 - Negative word correction.** Despite the numerous forms of negative words, their common goal is to introduce negation in the sentence. They were reduced to their basic negative form 'not'.
- POS tagging and lemmatisation—POS tagging by assigning words to their syntactic functions, such as noun, pronoun, adjective, verb and adverb, puts the words in context [35]; doing so allows lemmatising the word to the correct context, i.e., changing the word back to its basic form, which is an essential step in reducing word diversity and making recognition easier.

Since the features are extracted directly from social media messages, we used the BOW feature-extraction method. BOW is commonly used for

textual-based features [36], and works by decomposing the entire text into a group of singular words. In this study, we used it in combination with the n-gram technique to capture both singular words and word combinations that convey one meaning. Regarding word combinations, BOW checks the existence of a contiguous sequence of n words from the given text sample. In this study, we limited the checking to trigrams, since sequences of more than three words are infrequent in real-world texts. After decomposing the text as discussed above, the terms were sorted based on their frequency across the given dataset. A sample screenshot showing how the features are extracted using BOW is shown in Fig. 3. Whereas the screenshot is for an example scenario where the length of features is only 25 unigram words, 5000 unigram to trigram phrases were used for the length of features in our experiments. These settings were determined based on previous studies related to textual-based featurizing [16,17].

It is worth noting that one of the datasets used in our study is imbalanced, and this could affect prediction [37]. Hence, we also proposed a dynamic sampling process designed to optionally implement dynamic random over or under-sampling to overcome the issue of imbalanced data.

The next step is then to train the ML classifiers for the detection of mental depression. Towards this end, we implemented and tested several single and ensemble classifiers that provided excellent text recognition and detection performance in previous studies [16–18]. Those classifiers are:

- Logistic Regression:** A generalised linear model, it was originally developed by Nelder and Wedderburn [38] and further improved by Hastie and Tibshirani [39]. Generalised linear models use non-normal dependent variables and thereby overcome a major limitation of linear models, i.e., continuous and normally distributed

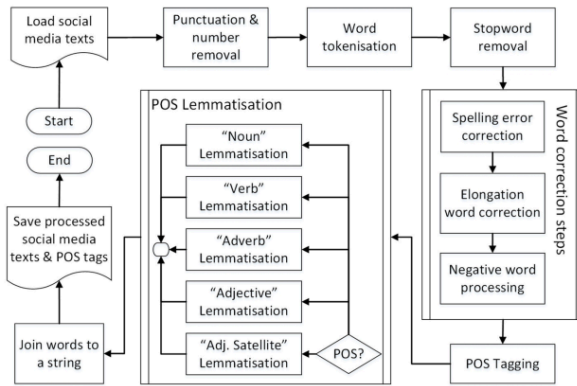


Fig. 2. Preprocessing steps [16].

dependent variables [40,41]. The dependent variables in LR are either unordered or ordered polytomous, and the independent predictor variables are either interval/ratio or dummy variables [42].

2. **Support Vector Machine:** The SVM is a supervised learning model that classifies new data after learning from training data [43]. It works by separating different classes with a hyperplane, and then attempts to maximise the separation distance from the hyperplane. The larger the distance, the lower the error generated by the classifier [44]. In this study, the SVM is used together with the linear kernel (LSVM), which is generally recommended for text classification [45].

3. **Multilayer Perceptron:** A feedforward artificial neural network that uses supervised learning [46,47], the MLP continually computes and updates all the weights in its network to minimise error. In its first phase, called the feedforward phase, the training data is forwarded to the output layer, following which the difference between this output and the desired target (the error) is backpropagated to update the weights of the network in the second phase [48]. In this study, we

utilised the Adam optimiser [49] to increase the performance of the MLP.

4. **Decision Tree:** The DT was developed by Quinlan [50] based on Hunt's algorithm [51], and is a useful tool for exploring the cause-and-effect chain. It builds a tree-like decision model for classification and prediction purposes and is typically used as a base classifier for ensemble models (e.g., BP, RF, and AB).
5. **Random Forest:** The RF is an ensemble of DT predictors where all decision trees are trained independently using random vectors. The strength of trees and their correlation determine the error generalisation of the RF. It is relatively robust to outliers and noise [52].
6. **Adaptive Boosting:** This ensemble combines many weak classifiers iteratively over several rounds [53]. Starting with equal weighting for all training data, weights of misclassified training data points are boosted, and a new classifier is created with the new, unequal weightings. This procedure is then repeated for a set of classifiers [54].

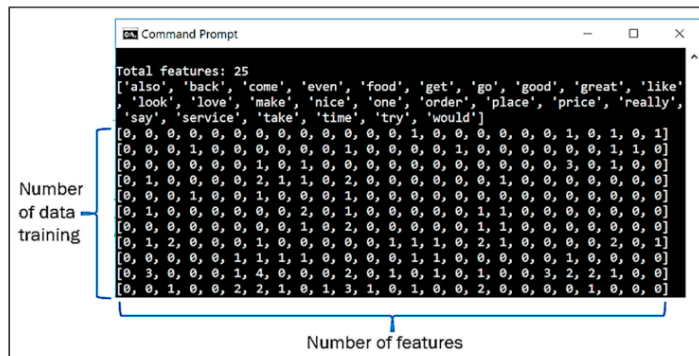


Fig. 3. A screenshot showing the features extracted using BOW for a simple hypothetical scenario with 25 unigram words used for the feature length.

7. **Bagging Predictors.** This ensemble builds a cluster of several single predictors, which are trained in a bootstrapping process that replicates the training set. Classification is performed using plurality voting [55]. The default BP implementation from scikit-learn, which is used in this study, uses the DT as its base predictor.
8. **Gradient Boosting.** GB, which comprises gradient boosted regression trees, provides a robust, competitive and interpretable algorithm for classification and regression. For binary classification, only a single regression tree is used [56].

A flowchart for the sampling process is shown in Fig. 4, and the overall design can be seen in Fig. 5.

We can see from Fig. 5 that all experiments were conducted using the N-fold Cross Validation (CV) method. CV is primarily used in classification and regression models to reduce the bias between an entire dataset and the training/testing sets [57], and it also helps avoid overfitting [58]. In CV, data is split into n disjoint folds or partitions, where $n-1$ folds (subsamples) are used for training and the remaining fold is used for testing. In this study, we set $n = 10$; i.e., all experiments were conducted using 10-fold CV. The overall measurements were calculated by averaging the results of each process [59,60].

Four well-known measurements—accuracy, precision, recall and F1 (or F-measure)—were used to evaluate the performance of the classifiers in terms of depression detection. Measurement components from scikit-learn [61] were implemented, and the relevant equations and functions are provided in Table 2. For detailed accuracy measurements, we calculated only the class accuracy, i.e., total correct detections of depression class testing samples against total depression class testing samples. Please note that since depression detection is a binary-target problem, scores for the detailed accuracy of positive class (depression) and binary recall will be the same.

4. Datasets

We primarily used datasets from Shen et al. [11] and Eye [12] in our experiments. These datasets were used to train and test the ML models using 10-fold CV. The texts in these two datasets were gathered from Twitter and automatically labelled as “Depression” or “Non-Depression”.

The dataset by Shen et al. [11] was constructed with the additional restriction that a record would be labelled as “Depression” only if its anchor tweets satisfied the strict pattern “(I m/I was/I am/I’ve been) diagnosed with depression”; the record would be labelled as “Non-Depression” if the user had never posted any tweet containing the character string “depress”. It should be noted that the dataset by Shen et al. [11] also contains an additional third set labelled as “Depression-candidate”—consisting of tweets that do not meet all the criteria to be labelled “Depression”—that was intended to boost the “Depression”

group. However, this third set was not used in our study because the sizes of the depression (first set) and non-depression (second set) classes are almost equal in this dataset. Including the third set in the “Depression” class would make the dataset heavily imbalanced. The dataset by Eye [12], on the other hand, only seeks the word ‘depression’ in the tweets. If the tweet contains the word ‘depression’, it is labelled as “Depression”, and “Non-Depression” otherwise. This dataset is highly imbalanced: depression class records are only 22% of the total records in this dataset.

To investigate whether the models trained with the above datasets could be successfully applied directly to other social media texts—besides the testing conducted with different subsets of the same dataset in 10-fold CV—we also ran the trained models on three non-Twitter based depression-class-only datasets [13–15]. These datasets could not be used to train the models, since they contain only depression class samples, but at least two classes (binary) are required to train the classifiers. The first dataset, by Tanwar [13], was constructed from the diary of a 17-year old girl, Victoria, who committed suicide because of depression. In her diary, Victoria recorded her feelings from during her depression until her suicide. The second dataset, by Komati [14], was constructed from Reddit posts about depression and suicide; in this study, we employed only a part of the depression section (50000 randomly selected records from a total of 348723 records). The third dataset was constructed by Virahonda [15] from Facebook posts. Studying the performance of our models on the above datasets allowed us to test whether they are generalised enough to be applied for detecting depression by analysing texts from diverse sources. Detailed information of all the datasets used in this study can be found in Table 3.

5. Results and discussion

During preprocessing, components from the Natural Language Toolkit [62] and Peter Norvig’s code for spelling correction [34] were applied. Similarly, components from scikit-learn were used to implement all the ML classifiers and to measure their performance (accuracy, precision, recall, and F1) [37]. In this section, we discuss the results of applying the proposed approach to depression detection using social media texts. We take depression as the positive class and non-depression as the negative class. All experiments were conducted using the 10-fold CV method, and each experiment was repeated 10 times.

5.1. Preliminary experiments on the proposed approach

Once the preprocessing and featurizing steps had been applied, preliminary experiments were run using the three best single classifiers LR, LSVM and MLP from a previous study [17] as well as the DT model. The results can be found in Table 4. These results in Table 4 are excellent—almost perfect—and, therefore, too good to be true. We suspected

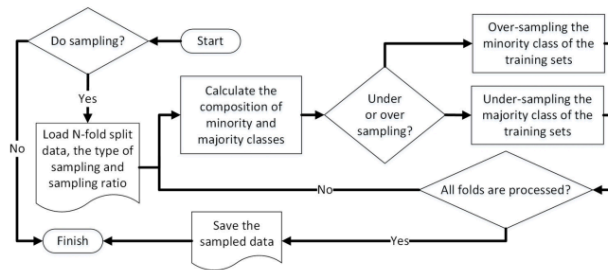


Fig. 4. The sampling process.

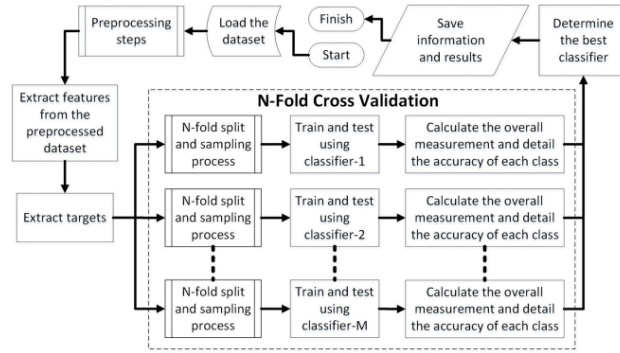


Fig. 5. Overall design.

Table 2
Measurement functions and formulas.

No	Name	Function	Equation
1	Accuracy	accuracy_score()	$A(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{k=0}^{n_{\text{classes}}-1} 1(\hat{y}_k = y_k)$ <p>where y is the set of predicted pairs, \hat{y} is the set of true pairs, and n_{samples} is the total number of samples.</p>
2	Precision	precision_score()	$P(y_k, \hat{y}_k) = \frac{tp}{tp + fp}$ <p>where k is the set of classes, y_k is the subset of y with class k, tp is true positive, and fp is false positive.</p>
3	Recall	recall_score()	$R(y_k, \hat{y}_k) = \frac{tp}{tp + fn}$ <p>where fn is false negative.</p>
4	F-measure/F1	f1_score()	$F_1(y_k, \hat{y}_k) = 2 * \frac{P(y_k, \hat{y}_k) * R(y_k, \hat{y}_k)}{P(y_k, \hat{y}_k) + R(y_k, \hat{y}_k)}$

Table 3
Datasets used in this study.

Dataset	Source	Total records	Depression records		Non-Depression records	
			Total	%	Total	%
Shen et al. [11]	Twitter	11877	6493	54.67	5384	45.33
Eye [12]	Twitter	10314	2314	22.44	8000	77.56
Tanwar [13]	Victoria's diary	62	62	100	0	0
Komati* [14]	Reddit about depression	50000	50000	100	0	0
Virahonda [15]	Facebook	9178	9178	100	0	0

(*) randomly selected from 348723 records.

the possibility of overfitting, i.e., the models performed well only on the datasets they had been trained on, but would not perform as well on other datasets. Thus, we conducted further experiments to test the performance of the trained models on three depression-class-only datasets listed in Table 3 (datasets by Tanwar, Komati, and Virahonda). The results of these experiments can also be seen in Table 4 ("Accuracy on Depression-class-only Datasets" column). As predicted, the models performed poorly on the depression-class-only datasets. One of the worst results was from the LR model trained using Shen et al.'s dataset when it was applied to Tanwar's dataset (0% accuracy). Even the best result (40.54% accuracy)—provided by LR trained using Eye's dataset and applied to Komati's dataset—is below 50%.

Since the single classifiers used in the above experiments had provided excellent results in previous studies on text detection [16,17], we suspected that the datasets used for training the models might be

responsible for this problem. A simple word frequency check revealed that the words 'diagnose' and 'depression' always exist in all the depression records of Shen et al.'s preprocessed dataset. This is expected, since the depression records in Shen et al.'s dataset were gathered using the strict pattern "(I'm/I was/I am/I've been) diagnosed with depression". Similarly, the depression records in Eye's dataset always contain the word 'depression'. As a result, the models learned to separate depression and non-depression classes based mainly on the existence of words such as 'diagnose' and 'depression' in the case of Shen et al.'s dataset and the word 'depression' in the case of Eye's dataset. Consequently, the models could not understand more subtle depression/non-depression patterns hidden in the text. Therefore, we deleted the word 'diagnose' from Shen et al.'s preprocessed dataset and the word 'depression' from Shen et al.'s and Eye's preprocessed datasets. Following this, we conducted another set of experiments on these

Table 4
Results of the preliminary experiments.

Dataset	Classifier	Overall Measurement (%)				Detailed Accuracy (%)		Accuracy on Depression-class-only Datasets (%)		
		Acc	Prec	Rec ^c	F1	Dep ^{a,f}	Non-Dep ^a	Tan ^b	Koma ^b	Vira ^b
Eye	LR	99.80	100	99.09	99.54	99.09	100	1.61	40.54	16.30
	LSVM	99.78	99.87	99.14	99.50	99.14	99.96	1.77	30.32	18.20
	MLP	99.12	99.78	96.27	97.99	96.27	99.94	9.19	31.32	31.24
	DT	99.74	99.57	99.26	99.41	99.26	99.88	13.39	48.59	31.35
Shen et al.	LR	99.77	100	99.58	99.79	99.58	100	0	16.76	13.86
	LSVM	99.80	99.94	99.68	99.81	99.68	99.93	0.32	12.09	13.67
	MLP	99.57	99.94	99.26	99.60	99.26	99.93	0	4.73	13.35
	DT	99.71	99.96	99.51	99.73	99.51	99.95	6.45	49.86	25.86

^a Dep = Depression; Non-Dep = Non-Depression.

^b Tan = Tanwar's dataset; Koma = Komati's dataset; Vira = Virahonda's dataset; these three datasets are single-class datasets comprised only of depression records.

^c Even though binary recall and accuracy in detecting positive (depression) classes will have the same values, both are presented here to show differences in the accuracy for depression and non-depression classes.

modified datasets, and the results can be seen in Table 5.

A comparison of the results in Tables 4 and 5 reveals a decrease in the overall measurements of all classifiers. The deterioration of the results is especially noticeable for the accuracy of depression record detection, which decreased by at least 13.17% for Shen et al.'s dataset and 21.25% for Eye's dataset. However, the detection accuracy significantly improved when the trained models were run on the other three depression-class-only datasets (Tanwar's, Komati's and Virahonda's datasets), with the improvement ranging from 27.42% to 67.58%. It should be noted that the models were trained using datasets with both depression and non-depression classes, however, they were also being tested on three datasets with depression only classes.

These results indicate that the models trained using the modified datasets are more general and can better detect depression in the text from datasets that they were not trained on. Table 5 also reveals that LR provided the best overall accuracy for both Shen et al.'s and Eye's datasets; however, the results were more diverse for the remaining datasets. Similarly, results show that Shen et al.'s dataset, whose construction was stricter than Eye's dataset, is better for depression detection on similar data, while it is less useful for the creation of a more general detection model. It should be noted that all the classifiers that were trained using our featurizing approach on Shen et al.'s dataset achieved higher scores than the baseline results in Shen et al. (85%) [11]. We also tested the differences in accuracies between each classifier for both datasets using the Wilcoxon signed-rank test; these results are shown in Table 6. It can be seen in Table 6 that the p-values of all pairwise comparisons are smaller than the significance level of 0.05, implying that all results are significantly different from each other.

5.2. Sampling or not sampling

As can be seen in Table 3, Eye's dataset is heavily imbalanced, i.e., there are many more depression records than non-depression records. We conducted experiments using the proposed dynamic sampling approach on Eye's dataset to test the effectiveness of the sampling procedure, and on Shen et al.'s dataset to check the effect of applying the same procedure on a slightly imbalanced dataset. The results of these experiments can be seen in Fig. 6 and Fig. 7.

It can be seen in Fig. 6 that both over and under-sampling improved the detection of the depression class. The rates of improvement in the case of over-sampling were 0.6% with the MLP, 1.5% with the LSVM, 8.7% with the LR, and 4.6% with the DT; whereas, for under-sampling, they were 8% with the MLP, 6.4% with the LSVM, 10.4% with the LR, and 12.5% with the DT. The detection of the non-depression class, however, worsened. From these results, we can conclude that the dynamic sampling procedure can increase the detection of depression class only when the dataset is imbalanced. In contrast, as can be seen in Fig. 7, dynamic sampling does not have any effect on a slightly imbalanced dataset.

The effects of dynamic sampling when models trained with Eye's and Shen et al.'s datasets were used for detecting depression in the depression-class-only datasets are shown in Fig. 8 and Fig. 9, respectively. The middle bars (i.e., no sampling) correspond to the scores in Table 5.

We can see in Fig. 8 that all the models trained using Eye's dataset improved the accuracy when combined with under-sampling, as opposed to when dynamic sampling was not applied. However, this effect was not always replicated in the case of over-sampling; only the LR

Table 5
Experimental results after the words 'depression' and 'diagnose' were deleted from the datasets.

Dataset	Classifier	Overall Measurement (%)				Detailed Accuracy (%)		Accuracy on Depression-class-only Datasets (%)		
		Acc	Prec	Rec ^c	F1	Dep ^{a,f}	Non-Dep ^a	Tan ^b	Koma ^b	Vira ^b
Eye	LR	92.61	93.32	72.21	81.38	72.21	98.51	69.19	90.40	69.65
	LSVM	91.57	84.84	76.04	80.17	76.04	96.07	60.00	85.92	70.44
	MLP	89.56	77.77	75.02	76.29	75.02	93.77	53.39	75.62	68.13
	DT	86.00	68.24	70.52	69.31	70.52	90.47	59.84	75.26	66.32
Shen et al.	LR	88.62	92.63	86.03	89.20	86.03	91.75	27.42	46.17	66.43
	LSVM	87.38	90.01	86.51	88.22	86.51	88.43	30.65	51.73	68.99
	MLP	85.13	86.65	86.07	86.35	86.07	84.01	28.39	52.70	64.30
	DT	82.23	83.46	84.20	83.82	84.20	79.86	47.90	52.62	66.93

^a Dep = Depression; Non-Dep = Non-Depression.

^b Tan = Tanwar's dataset; Koma = Komati's dataset; Vira = Virahonda's dataset; these three datasets are single-class datasets comprised only of depression records.

^c Even though binary recall and accuracy in detecting positive (depression) classes will have the same values, both are presented here to show differences in the accuracy for depression and non-depression classes.

Table 6

Wilcoxon signed-rank test results (a. Eye's dataset; b. Shen et al.'s dataset).

a. Eye's dataset				b. Shen et al.'s dataset			
	LSVM	MLP	DT		LSVM	MLP	DT
LR	0.00028	<0.00001	<0.00001	LR	0.00026	<0.00001	<0.00001
LSVM		<0.00001	<0.00001	LSVM		<0.00001	<0.00001
MLP			<0.00001	MLP			<0.00001

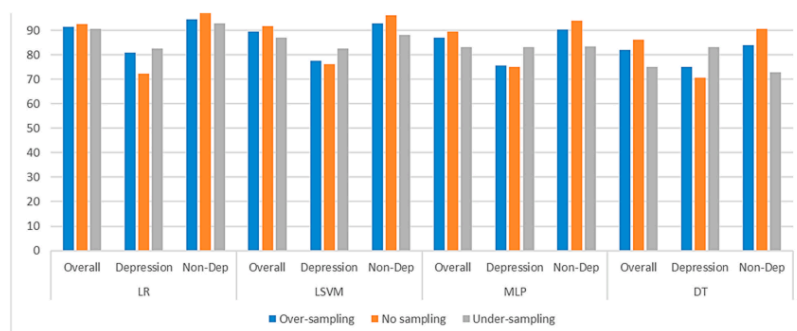


Fig. 6. Accuracy of sampling effect on Eye's dataset (heavily imbalanced).

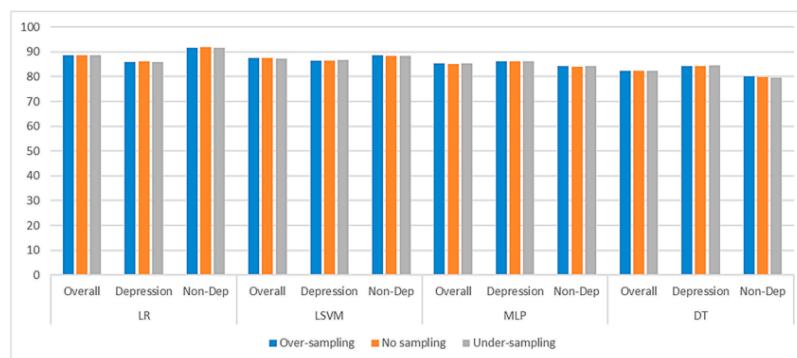


Fig. 7. Accuracy of sampling effect on Shen et al.'s dataset (slightly imbalanced).

model improved accuracy with over-sampling compared to no sampling. We suspect that, with over-sampling, the training records are duplicated randomly in the case of the smaller class (the depression class in our case), which increases the population of the smaller class but not its diversity. This condition makes the trained models more exclusive to the trained dataset, but not the other datasets. In the case of models trained with Shen et al.'s dataset (see Fig. 9), as can be predicted from the results in Fig. 7, dynamic sampling did not have a significant impact on the detection accuracy. However, surprisingly, both sampling methods significantly improved the detection accuracy in the case of Komati's depression dataset with the LR and LSVM models. Based on the results in

Figs. 8 and 9, we can conclude that LR performed the best with both sampling methods, and that dynamic under-sampling increased the accuracy of depression detection.

5.3. Experiments on ensemble classifiers

Our last set of experiments was conducted to test the performance of our method on several ensemble models, namely the AB, BP, GB, and RF. These ensemble models had performed well in previous studies on text analysis (e.g. see [17]). Based on the results in the previous section, these experiments were carried out together with dynamic

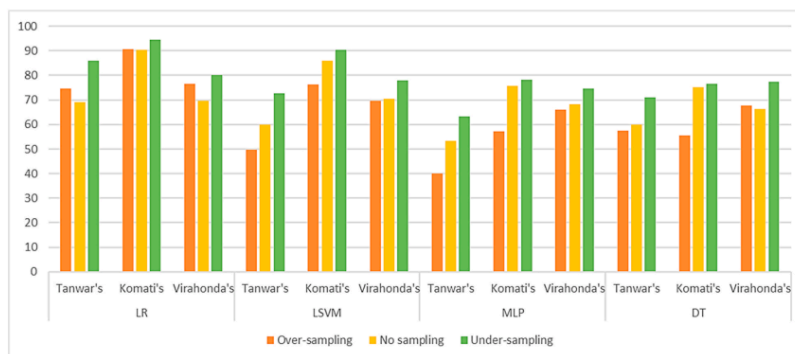


Fig. 8. Accuracy of sampling effect of Eye's trained models on depression-class-only datasets.

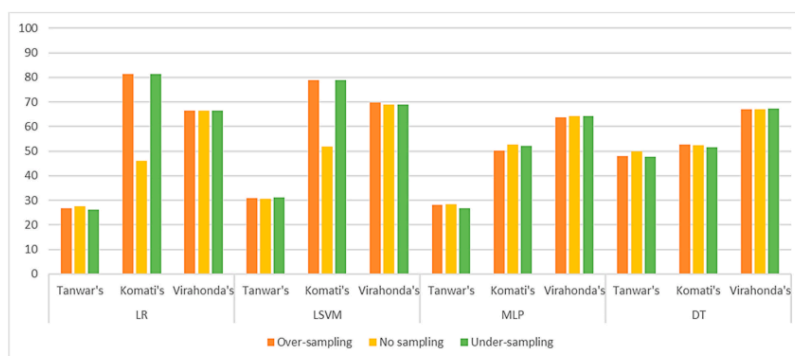


Fig. 9. Accuracy of sampling effect of Shen et al.'s trained models on depression-class-only datasets.

under-sampling. The results of the experiments on Eye's and Shen et al.'s datasets and the depression-class-only datasets can be seen in Fig. 10 and Fig. 11, respectively.

The results in Fig. 10 indicate that, among the ensemble models, the RF provided the most balanced accuracy for both depression and non-depression classes in both Eye's and Shen et al.'s datasets. However, these results were not better than those provided by the single classifier LR. While we are unable to pinpoint the reason why the results were not better than LR (numerous factors could influence the results of an ensemble model), we believe that the base classifiers used by the ensemble models might be significantly impacting their results. As discussed above, the BP, AB and RF all use the DT model as their base classifier. Results in Table 5 show the DT to be the worst classifier among those tested, and the performance of DT-based ensemble models was similar to the DT itself. Regarding the performance of GB, which uses a single regression tree for classification/prediction in the vein of LR, we suspect the boosting process in GB is less suitable for depression detection than LR. On the other hand, when tested on the three depression-class-only datasets (see the results in Fig. 11), the RF provided better

results than all single and ensemble classifiers. Therefore, we can conclude that the RF model is better at depression detection using general texts than other classifiers.

6. Conclusion

Depression is the most prevalent mental disorder and the main cause behind more than two-thirds of suicides every year. Unfortunately, many cases go untreated because of failure to detect or self-denial. Several studies agreed that, given the exponential increase in social media usage, social media messages can be used as a valuable source for monitoring several mental health issues, including depression.

In this paper, we demonstrated that our generalised approach using ML methods and social media texts can be effectively used to detect signs of depression. Our ML models proved effective even when trained with texts that did not contain the words 'depression' or 'diagnosis'—social media messages by those suffering from depression rarely include such words. It is important to note that the approach presented in this paper performed well even when tested on datasets that were unrelated to the

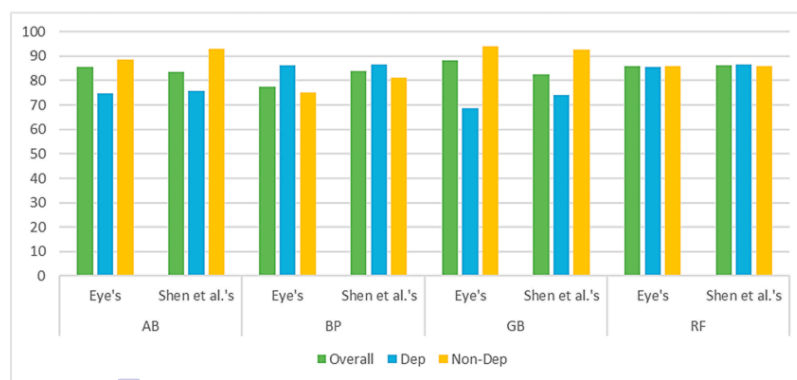


Fig. 10. Accuracy of several ensemble classifiers with under-sampling for Eye's and Shen et al.'s datasets.

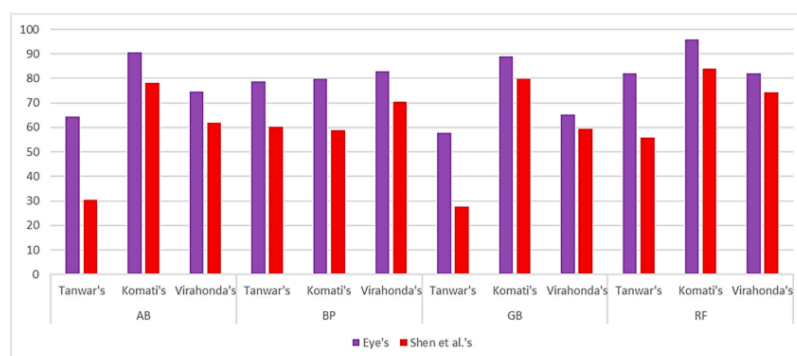


Fig. 11. Accuracy of ensemble models that are trained using Eye's and Shen et al.'s datasets but tested on Tanwar's, Komati's and Virahonda's datasets.

training datasets. This contrasts with most studies in the literature that use portions of the same dataset for training and testing. Our results also indicated that using less strictly constructed datasets can be more beneficial than more strictly constructed datasets, especially when the models would be used for detecting depression in messages from a variety of sources.

Finally, it should be noted that the approach presented in this paper uses supervised ML classifiers, and therefore, the approach is limited to using labelled datasets for training the classifiers. Overcoming this limitation by including unsupervised classifiers is a potential future research area.

An additional observation is that, in a heavily imbalanced dataset, dynamic sampling could increase the accuracy of the less populous class but, at the same time, decrease the accuracy of the more populous class. However, this can be beneficial if the goal is to detect depression and the less populous class is the depression class (which is the case in this study).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the University of Newcastle's College Multidisciplinary Strategic Investment Funding for 2021.

References

- [1] A. Priya, S. Garg, N.P. Tigga, Predicting anxiety, depression and stress in modern life using machine learning algorithms, *Procedia Computer Science* 167 (2020) 1258–1267.
- [2] N.S. Srimadhar, S. Lalitha, An end-to-end model for detection and assessment of depression levels using speech, *Procedia Computer Science* 171 (2020) 12–21.
- [3] H.S. Alsagari, M. Ykhlef, Machine learning-based approach for depression detection in Twitter using content and activity features, *IEEE Transactions on Information*

- and Systems E103.D (8) (2020) 1825–1832, <https://doi.org/10.1587/transinf.2020EDP7023>.
- [4] S. Nikolin, Y.Y. Tan, A. Schwaab, A. Moffa, C.K. Loo, D. Martin, An investigation of working memory deficits in depression using the n-back task: a systematic review and meta-analysis, *J. Affect. Disord.* 284 (2021) 1–8, <https://doi.org/10.1016/j.jad.2021.01.084>.
 - [5] S. Yadav, T. Kalin, S. Gupta, U. Bhatti, P. Priyadarshi, Predicting depression from routine survey data using machine learning, in: Proceedings of the 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp. 163–168.
 - [6] A.U. Hassan, J. Hussain, M. Hussain, M. Sadiq, S. Lee, Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression, in: Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, South Korea, 2017, pp. 138–140.
 - [7] O. Edo-Osagie, B. De La Iglesia, I. Lake, O. Edeghere, A scoping review of the use of Twitter for public health research, *Comput. Biol. Med.* 122 (2020) 103770, <https://doi.org/10.1016/j.combiomed.2020.103770>.
 - [8] V. Mishra, T. Gang, A systematic study on predicting depression using text analytic, *J. Fund. Appl. Sci.* 10 (2) (2018) 293–307, <https://doi.org/10.4314/jfas.v10i2.21>.
 - [9] J. Kim, J. Lee, E. Park, J. Han, A deep learning model for detecting mental illness from user content on social media, *Sci. Rep.* 10 (1) (2020) 11846, <https://doi.org/10.1038/s41598-020-68764-y>.
 - [10] K. Mahnken, Survey: more young people are depressed during the pandemic. But they may be using social media to cope. The 74 million, <https://www.the74million.org/survey-more-young-people-are-depressed-during-the-pandemic-but-they-may-be-using-social-media-to-cope/>, 2021. Accessed 28 April 2021.
 - [11] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, W. Zhu, Depression detection via harvesting social media: a multimodal dictionary learning solution, in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017, pp. 3838–3844.
 - [12] B.B. Eye, Depression Analysis, 1 edn, Kaggle, 2020.
 - [13] R. Tanwar, Victoria Suicide Data, Kaggle, 2020.
 - [14] N. Kamal, y/SuicideWatch and R/depression Posts from Reddit, Kaggle, 2020.
 - [15] S. Virabanda, Depression and Anxiety Comments, 1 edn, Kaggle, 2020.
 - [16] G.S. Budhi, R. Chiong, Z. Wang, Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features, *Multimed. Tool. Appl.* 80 (9) (2021) 13079–13097, <https://doi.org/10.1007/s11042-020-10299-5>.
 - [17] G.S. Budhi, R. Chiong, I. Pranata, Z. Hu, Using machine learning to predict the sentiment of online reviews: a new framework for comparative analysis, *Arch. Comput. Methods Eng.* 28 (2021) 2543–2566, <https://doi.org/10.1007/s11831-020-09464-8>.
 - [18] G.S. Budhi, R. Chiong, I. Pranata, Z. Hu, Predicting rating polarity through automatic classification of review texts, in: Proceedings of the 2017 IEEE Conference on Big Data and Analytics (ICBDA), Kuching, Malaysia, 2017, pp. 19–24.
 - [19] J.R. Sato, J. Moll, S. Green, J.F. Doakin, C.E. Thomaz, R. Zahn, Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression, *Psychiatr. Res.* 233 (2) (2015) 289–291, <https://doi.org/10.1016/j.psychres.2015.07.001>.
 - [20] P. Kumar, S. Gang, A. Garg, Assessment of anxiety, depression and stress using machine learning models, *Procedia Computer Science* 171 (2020) 1989–1998, <https://doi.org/10.1016/j.procs.2020.04.213>.
 - [21] B. Ojeme, A. Mbogho, Selecting learning algorithms for simultaneous identification of depression and comorbid disorders, *Procedia Computer Science* 96 (2016) 1294–1303, <https://doi.org/10.1016/j.procs.2016.08.174>.
 - [22] J. Gratch, R. Artstein, G. Lucas, G. Stratos, S. Scherer, A. Nazarian, R. Wood, J. Boberg, D. DeVault, S. Marsella, D. Traum, S. Rizzo, L.-P. Morency, The distress analysis interview corpus of human and computer interviews, in: Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC), Reykjavik, Iceland, 2014, pp. 3123–3128.
 - [23] A. Samareh, Y. Jin, Z. Wang, X. Chang, S. Huang, Detect depression from communication: how computer vision, signal processing, and sentiment analysis join forces, *IJSE Transactions on Healthcare Systems Engineering* 8 (3) (2018) 196–208, <https://doi.org/10.1080/2472579.2018.1496494>.
 - [24] Y. Chen, B. Zhou, W. Zhang, W. Gong, G. Sun, Sentiment analysis based on deep learning and its application in screening for perinatal depression, in: Proceedings of the IEEE 3rd International Conference on Data Science in Cyberspace (DS-Cy), Guangzhou, China, 2018, pp. 451–456.
 - [25] M.R. Islam, M.A. Kabir, A. Ahmed, A.R.M. Kamal, H. Wang, A. Ulhaq, Depression detection from social network data using machine learning techniques, *Health Inf. Sci. Syst.* 6 (1) (2018) 8, <https://doi.org/10.1007/s13755-018-0046-0>.
 - [26] S.G. Burduso, M. Errecalde, M. Montes-y-Gómez, A text classification framework for simple and effective early depression detection over social media streams, *Expert Syst. Appl.* 133 (2019) 182–197, <https://doi.org/10.1016/j.eswa.2019.05.023>.
 - [27] I. Fatima, B.U.D. Abbasi, S. Khan, M. Al-Saeed, H.F. Ahmad, R. Mumtaz, Prediction of postpartum depression using machine learning techniques from social media text, *Expert Syst.* 36 (4) (2019) e12409, <https://doi.org/10.1111/exsy.12409>.
 - [28] R. Chiong, G.S. Budhi, S. Dhakal, Combining sentiment lexicons and content-based features for depression detection, *IEEE Intell. Syst.* 36 (6) (2021) in press.
 - [29] E.M.S. Filho, H.C. Veiga Rey, R.M. Frjatag, D.M. Arrowsmith Cook, L.N. Dalbonio de Carvalho, A.L. Pinho Ribeiro, J. Amaral, Can machine learning be useful as a screening tool for depression in primary care? *J. Psychiatr. Res.* 132 (2021) 1–6, <https://doi.org/10.1016/j.jpsychires.2020.09.025>.
 - [30] N. Jothi, W. Hussain, N.A. Rashid, Predicting generalized anxiety disorder among women using Shapley value, *J. Infect Public Health* 14 (1) (2020) 103–108, <https://doi.org/10.1016/j.jiph.2020.02.042>.
 - [31] H. Jung, H.A. Park, T.M. Song, Ontology-based approach to social data sentiment analysis: detection of adolescent depression signals, *J. Med. Internet Res.* 19 (7) (2017) e259, <https://doi.org/10.2196/jmir.7452>.
 - [32] B. Sutter, R. Chiong, G.S. Budhi, S. Dhakal, Predicting psychological distress from ecological factors: a machine learning approach, in: Proceedings of the 34th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE), Kuala Lumpur, Malaysia, 2021, pp. 1–12.
 - [33] C. Lin, P. Hu, H. Su, S. Li, J. Mei, J. Zhou, H. Leung, SenseMood: depression detection on social media, in: Proceedings of the 2020 International Conference on Multimedia Retrieval, Dublin, Ireland, 2020, pp. 407–411.
 - [34] P. Norvig, How to write a spelling corrector, 2016. <https://norvig.com/spell-corr-ect.html>. Accessed June 01 2018.
 - [35] W. Elaiwi, G. Naymat, The impact of applying different preprocessing steps on review spam detection, *Procedia Computer Science* 113 (2017) 273–279.
 - [36] X. Deng, Y. Li, J. Weng, J. Zhang, Feature selection for text classification: a review, *Multimed. Tool. Appl.* 78 (3) (2019) 3797–3816, <https://doi.org/10.1007/s11042-018-6083-5>.
 - [37] Z. Hu, R. Chiong, I. Pranata, Y. Bao, Y. Lin, Malicious web domain identification using online credibility and performance data by considering the class imbalance issue, *Ind. Manag. Data Syst.* 119 (3) (2019) 676–696, <https://doi.org/10.1108/IMDS-02-2018-0072>.
 - [38] J.A. Nelder, R.W.M. Wedderburn, Generalized linear models, *J. Roy. Stat. Soc.* 135 (3) (1972) 370–384, <https://doi.org/10.2307/234614>.
 - [39] T. Hastie, R. Tibshirani, Generalized Additive Models, Chapman and Hall/CRC, United Kingdom, 1990.
 - [40] G.H. Dumtman, M.-H.R. Ho, Generalized linear models, in: A.J. Dobson, A. G. Barnett (Eds.), *An Introduction to Generalized Linear Models*, SAGE, 2011, pp. 2–6.
 - [41] A.J. Dobson, A.G. Barnett, *An Introduction to Generalized Linear Models*, third ed., CRC Press, Boca Raton, 2008.
 - [42] S. Menard, Logistic Regression: From Introductory to Advanced Concepts and Applications, SAGE, Los Angeles, 2010.
 - [43] R. Chiong, Z. Fan, Z. Hu, F. Chiong, Using an improved relative error support vector machine for body fat prediction, *Comput. Methods Progr. Biomed.* 198 (2021) 105749, <https://doi.org/10.1016/j.cmpb.2020.105749>.
 - [44] C. Campbell, Y. Ying, Learning with Support Vector Machines, Morgan & Claypool, 2011.
 - [45] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, *ACM Transactions on Intelligent Systems and Technology* 2 (3) (2011) 1–27, <https://doi.org/10.1145/1961189.1961199>.
 - [46] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, *Heliyon* 4 (11) (2018), e00938, <https://doi.org/10.1016/j.heliyon.2018.e00938>.
 - [47] O.I. Abiodun, M.U. Kiru, A. Jantan, A.E. Omolara, K.V. Dada, A.M. Umar, O. U. Linus, H. Arshad, A.A. Kazaure, U. Gana, Comprehensive review of artificial neural network applications to pattern recognition, *IEEE Access* 7 (2019) 158820–158846, <https://doi.org/10.1109/access.2019.2945545>.
 - [48] D.E. Rumelhart, J.L. McClelland, Learning internal representations by error propagation, in: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, 1987, pp. 318–362.
 - [49] D.P. Kingma, J.L. Ba, Adam: A method for stochastic optimization, in: Proceedings of International Conference on Learning Representations, San Diego, USA, 2015, pp. 1–15.
 - [50] J.R. Quinlan, Induction of decision trees, *Mach. Learn.* 1 (1) (1986) 81–106, <https://doi.org/10.1007/bf00116251>.
 - [51] E.B. Hunt, J. Martin, P.J. Stone, Experiments in Induction, Academic Press, New York, 1966.
 - [52] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32, <https://doi.org/10.1023/a:1010933404324>.
 - [53] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, *J. Comput. Syst. Sci.* 55 (1) (1997) 119–139, <https://doi.org/10.1006/jcss.1997.1504>.
 - [54] J. Zhu, H. Zou, S. Rosset, T. Hastie, Multi-class AdaBoost, *Stat. Interface* 2 (2009) 349–360.
 - [55] L. Breiman, Bagging predictors, *Mach. Learn.* 24 (2) (1996) 123–140, <https://doi.org/10.1007/bf00058655>.
 - [56] J.H. Friedman, Greedy function approximation: a gradient boosting machine, *Ann. Stat.* 29 (5) (2001) 1189–1232.
 - [57] Q. Ren, M. Li, S. Han, Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives, *Big Earth Data* 3 (1) (2019) 8–25, <https://doi.org/10.1080/20964471.2019.1572452>.
 - [58] T. Xiong, Y. Bao, Z. Hu, R. Chiong, Forecasting interval time series using a fully complex-valued RBF neural network with DPSO and PSO algorithms, *Inf. Sci.* 305 (2015) 77–92, <https://doi.org/10.1016/j.ins.2015.01.029>.

- [59] R. Primartha, B. Adhi Tama, A. Arfiansyah, K. Januar Miraswan, Decision tree combined with PSO-based feature selection for sentiment analysis, *J. Phys. Conf.* 1196 (2019) 012018, <https://doi.org/10.1088/1742-6596/1196/1/012018>.
- [60] T.-L. Nguyen, S. Kavuri, M. Lee, S.O. Hwang, A fuzzy convolutional neural network for text sentiment analysis, *J. Intell. Fuzzy Syst.* 35 (6) (2018) 6025–6034, <https://doi.org/10.3233/jifs-169843>.
- [61] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (85) (2011) 2825–2830.
- [62] S. Bird, E. Klein, E. Loper, *Natural Language Processing with Python*, O'Reilly Media, Inc., USA, 2009.

A textual-based featuring approach for depression detection using machine learning classifiers and social media texts

ORIGINALITY REPORT

22%

SIMILARITY INDEX

17%

INTERNET SOURCES

10%

PUBLICATIONS

17%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Vikrant University

Student Paper

6%

2

pubmed.ncbi.nlm.nih.gov

Internet Source

3%

3

ieeecs-media.computer.org

Internet Source

2%

4

Submitted to Sheffield Hallam University

Student Paper

2%

5

link.springer.com

Internet Source

1%

6

ijetms.in

Internet Source

1%

7

Submitted to RMIT University

Student Paper

1%

8

eudl.eu

Internet Source

1%

9

Submitted to ESCP-EAP

Student Paper

1%

10

www.irjmets.com

Internet Source

1%

11

Submitted to Siksha 'O' Anusandhan University

Student Paper

1%

12

ijarsct.co.in

Internet Source

1%

13

aircconline.com

Internet Source

1%

14

dspace.uef.fi
Internet Source

1%

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On