International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Identification of Psychological Resilience over Social Media Using Stacked Ensemble Learning Algorithm

Tejaswita Garg*[1], Sanjay K. Gupta[2]

Submitted: 28/12/2023 **Revised**: 04/02/2024 **Accepted**: 12/02/2024

Abstract: Online communication over social media affecting psychological resilience that helps to pre-identify neurological disorder activities. Currently, existing research commonly uses a single model for such detection. This study suggests a stacked ensemble learning algorithm that ensembles four base classifiers including Support Vector, Random Forest, K Nearest Neighbor, Catboost, and a Meta classifier as Logistic Regression, along with a variety of word embeddings including Word2Vec, GloVe and FastText on the training corpus that is performed over Twitter public dataset to identify such neurological disorder problems among individuals. The training and testing models are tuned and then calculates the efficiency of proposed model in terms of metric calculation scores via Precision, Accuracy, Recall and F1-scores. The proposed ensemble model performed better over standalone models and results are then evaluated using confusion matrix & RoC curves. It also gives comparison based on execution time among all the classifiers. Hence, this research aimed to the earlier disclosure of such symptoms that can helps to increase psychological resilience and ultimately lowering the affect of mental hazard problems.

Keywords: *Psychological Resilience, Digital Footprint, Stacked Ensemble Learning, Word Embedding, Neurological Disorder, Sentiment Analysis.*

1. Introduction

The popularity of online social networking sites (SNSs), particularly among teenagers, is enormous. Digital footprint left by activity on these websites can be examined to learn more about the behavioural online correlates of teenage psychological distress and, ultimately, to enhance detection and intervention methods. A prevalent mental illness, depression is one of the leading factors in disability and suicide globally [1]. About 300 million people globally deal with depression, as reported by the World Health Organization and it continued rising at greater speed [2]. Likewise, 70% of sufferers are afraid to see a physician throughout the initial stages of their condition, despite the fact that earlier treatment for depression can lessen the disease's detrimental effects [3]. They are wary due to discrimination and stigma. Researchers have discovered that stigma and prejudice are widespread underlying issues, even though not every patient feels such issues [4]. Determining depression is therefore a difficult task.

Depression has emerged as a significant global health issue also during pandemic COVID19, affecting 322 million people globally [5]. Numerous chronic diseases, including diabetes, heart disease, and others, can develop among depressed people as a result of depression. It is the secondary main factor in the emergence of chronic illnesses [6] [7]. Serious depression disorder may lead to suicide thoughts or suicidal attempts.

The use of machine learning algorithms is to deduce meaningful information from data of many networks resulted as the tremendous advertisement in information and technology. Such algorithms are frequently employed in the medical fields and their application in the psychological field is still somewhat limited. Psychometrics and Psychological analysis have both employed statistical inferences for many years. Following the Cambridge Analytica Scandal, Machine learning has mostly gained attention for its application in psychometrics. Researchers are currently tuning to algorithms from statistical conclusions in psychological testing and analysis due to the replicability issues with statistical inference [8]. Psychological Resilience refers to the ability to cope with mental health. It's ultimately inversely proportional to the mental risks. Such earlier identification of mental problems that grows with stress, anxiety or depression can helps to increase such resilience when prompted at an earlier stage [9].

Identifying predictors of neurological disease has recently been done using machine learning techniques like SVM and Random forest [10], [11]. Among these, the stacked ensemble machine can lessen the danger of biases that an individual machine learning model can have by increasing accuracy by mixing two or more separate machine learning models with a meta-model. Additionally, it has been demonstrated that it is more accurate in predicting outcome

[1] Research Scholar, School of Studies in Computer Science & Applications, Jiwaji University, Gwalior, Madhya Pradesh, INDIA - 474011
ORCID ID : 0009-0005-5097-4326
[2] Professor & Head, School of Studies in Computer Science & Applications, Jiwaji University, GWALIOR, Madhya Pradesh, INDIA - 474011
ORCID ID : 0009-0003-1499-8641
** Corresponding Author Email: tejaswitagarg@email.com*

factors [12]. However, there are still insufficient research that used the stacked ensemble machine and patient data to forecast a disease.

Studies in the medical field have been trying to find potential risks for depressive disorders for the past ten years, utilizing both data mining approaches and conventional statistical analysis models (regression model) [13]. Recent studies of Byeon H. and Kandel I. et al. have employed the stacked ensemble technique in particular to get over the constraints of an individual machine learning approach [14], [15]. This technique estimates classes by merging various learning algorithms. Due to the adoption of a Meta model that uses the information predicted by separate algorithms and stacking ensemble forecasts, it's known that it has a greater precision than a particular machine learning technique. However, it was also shown that, irrespective of the nature of algorithms used in the base and Meta learner model, the accuracy of a stacked ensemble model was worse compared to a standalone machine learning model. In order to develop the best ensemble model that can predict depressive illnesses, a lot of stacked generalization ensemble machine learning theories must be conducted.

1.1. Research Questions

The following research questions can be imposed in this work:

- How social media helps to determine mental health disorders?
- Why ensemble learning better than standalone models for such prediction?
- How psychological resilience can be improved using machine learning approach?

The stacking ensemble was used in this preliminary study to examine the key variables that may predict depression in neurological disorder patients. Baseline data were also presented in order to create a nomogram predictive index for the future identification of high risk subgroups for depression among neurological disorder disease patients.

1.2. Contributions

The contribution of this study can be summed up as follows:

- This study proposed an Optimized Stacked Ensemble Learning Algorithm (Algorithm 1) utilizing stacked generalization approach for neurological disorder detection through digital footprints.
- The performance of the model is evaluated using three word embedding models over huge corpus dataset comprising 10308 Twitter posts to obtain pre-trained term vectors that have been trained on a large number of words and are able to capture word semantics.
- Extensive experimentation done using ensemble technique having four base classifiers as SVC, RFC, KNN and CBC; and a Meta classifier as LR with stratified 10-fold cross validation to get improved results on the basis of execution time taken by standalone and machine learning classifiers.
- The performance of the model is evaluated in terms of metric evaluation including Accuracy, Precision, Recall, F1-Score, Confusion matrix and RoC curves.

We assess that this study aimed to increase psychological resilience among individuals through earlier identification of neurological factors that helps to lowering the effect of mental risks over social media platforms.

The remaining sections of the paper are organized as follows: The review of existing methods and approaches is discussed in section 2.The proposed architecture and experimental setup is presented in section 3, which also includes a dataset preparation, the implementation procedure, and the classifiers. The experimental outcome, additional observations, and a comparative analysis are covered in Section 4. Section 5 presents conclusion and strategy for the future.

2. Background Study

In order to learn about the methods and approaches utilized in the previous works and identify any research gaps, this section has read through a number of related research articles.

Jayawickreme [16] determine that the Korea Republic users would soon experience a depression. The RFC was used to build the predictive model. SMOTE was utilized to address the problems of class disparities. The accuracy of this study was 86.20%. This study also showed that the most important elements influencing the beginning of depression are health and happiness with social and familial relationships.

It has been determined by Na [17], author focused on the senior groups. They have used 10 different classification techniques to predict depression among peoples. The Random Forest performed the best out of the ten classifiers.

From author's work in [18], they have used 284 older individuals' psychometric and demographic data to forecast the level of depression. For identifying the persistence of depression, they used the XGBoost method and evaluated its efficacy in comparison to LR model. They claimed that XGBoost outperformed LR in terms of performance.

The author [19] created a mobile-based application to foresee PPD. Using the socioeconomic, therapeutic and psychological data of 1397 new mums, they assessed the effectiveness of NB, ANN, LR, and SVM. They claimed that their study was the first reliable Clinical Decision Support (CDS) system capable of determining whether PPD existed in the first week following delivery.

In this study, authors [20] used the CNN and Unsupervised Extreme Learning Machine (US-ELM) algorithms to apply mass detection to breast imaging datasets. They calculated that their classification accuracy was the greatest at 86.5%. Our study is superior because in addition to skipping the feature extraction stage, it also meets all the requirements in addition to the benign and malignant classes.

This research study has published a comparative analysis of several heart disease prediction methods, including SVM, NB, LR , NN & voting classifiers. The voting classifier has a 90% accuracy rate, which is the highest. The author also suggested that a GUI be created [21].

After analyzing the experimental data [22], researchers came to the conclusion that the J48 tree technique is the best classifier for predicting heart disease since it is more accurate and takes less time to create. It is obvious that the J48 algorithm with lower error pruning and Logistic Model Tree (LMT) method have the highest accuracy.

In this study [23], author proposed a machine learning approach as Naive Bayes classification that gives the accuracy value of 76.6%.

Some detection systems use the participant's self-reported health state as the basis for their ground truth labeling. The majority of studies, including those by authors rely on self-reports of depression status [24]. Posts that are suggestive of depression are recognized and utilized as data for training via supervised learning algorithms. Unfortunately, the level of depression is never evaluated by a psychologist or a questionnaire when datasets are created in this way.

ALSAGRI & YKHLEF [25] gathered data from 500 Twitter user profiles. They used a number of classifiers, but SVM had the highest accuracy for detecting depression at 82.5%.

In their study [26], authors used machine learning (ML) and NLP (natural language processing) classifiers to predict sadness from Reddit forums. Support vector machine (SVM) provided them with 90% accuracy to determine from linked words of different emotions after they analyzed 1,293 depression suggestive posts and 548 ordinary posts.

According to the study in [27], depressed persons use Twitter as a tool for promoting social awareness while non-depressed individuals use it as a tool for information collecting. Another well-known social media platform is Reddit.

Young minds matter (YMM) provided a dataset that was compiled in this study [28]. They discovered 11 key signs, such as sadness, boredom, and irritability that can be used to identify depression in children and adolescents. Someone is sad if they exhibit any five of the eleven signs listed above. In just 315 milliseconds, RF was able to predict 99% of cases of depression.

The authors [29] used ML and NLP to detect depression in 1,335 references from various datasets. They made advantage of Reddit forums, Facebook, and Twitter.

This study used Reddit dataset and applied SVM, KNN and multimodal NB through VADER sentiment analysis and achieved 89.36% accuracy score with 10 fold CV for NB model [30].

This study suggests that the Natural language processing and explainable artificial intelligence (AI) are coupled to examine and evaluate depression-related linguistic biomarkers for English-Urdu textual posts [31].

This study compare model performance with three different conventional classifiers on this classification task by combining psycholinguistic information in a rule-based estimator with 82% accuracy score on 9210 Reddit posts [32].

The authors of this study compared the linguistic content of people in online forums for various forms of mental distress using the Linguistic Inquiry and Word Count program (LIWC) [33].

This research used multimodal psychological highly imbalanced dataset, WESAD over 12 subjects and worn over chest and wrist to evaluate over three classes stress, neutral and amusement. Among various machine learning algorithms, RF performed best with accuracy of 84.17 [34].

Table 1 gives a summary of the surveyed deep learning-based depression detection methods and illustrates comparative analysis for mentioned research. This table clearly shows how heavily text data is used. Recently, we've noticed a shift away from manually created features and towards intricate neural word embedding models. This follows a general trend in data science where robust text embedding models have become the state-of-the-art. Additionally, some research on depression has been done using the Reddit community [35-37]. Reddit's forum appealed to researchers because it permits lengthy submissions with no word count restrictions [38]. Some researchers are only interested in posts that are both suicide-indicative and non-suicidal after characterizing the deeper relationships between language and sadness.

Therefore, in the existing literature we found that stacked ensemble learning is not that much explored. Amongst all, twitter dataset has been chosen for performing social media analytics. Hence, this paper proposed a stacked ensemble learning approach for better accuracy over existing literature.

Table 1. Summary of findings for varied dataset sources with numerous computational approaches.

Year	Research	Data Users	Dataset Source	Method	Comparative Analysis	Metric
2020	Na et al.	6,588	Korea Users	Random Forest	Individual RF model evaluated on Korean user's dataset to predict depressive disorder.	86.20
2017	Sau and Bhakta	520	Geriatric patient	Random Forest	Results were compiled on ten different single ML techniques.	89
2019	Hatton et al.	284	Geriatric patient	XGBoost	XGBoost outperformed over other classifiers.	74
2015	Jiménez-Serrano et al.	1,397	Postpartum period mums	Naive Bayes	Postpartum period mums disorder (PPD) detected using various classifiers.	79
2019	Wang and Li	400	Mammograms breast images	CNN and ELM	Deep learning architecture has been used to classify breast imaging datasets.	86.5
2020	Sri	303	Cleveland heart	Voting Ensemble	Individual ML algorithms applied to classify heart disease datasets.	90
2017	Jaymin Patel	303	Cleveland heart	J48, RF and LR	Predicting heart disease using J48 and LMT learning classifiers.	56.76
2021	Rinki Chhaterjee	7,146	Facebook	Naive Bayes	Limited to single ML classifier.	76.6
2018	Pirina and öltekin	800	Reddit with eight subsets	Linear SVM	Limited to only the outcomes of linear classifiers with straightforward character and word bag-of-n-gram features were presented.	98.2
2020	Alsagri and Ykhlef	500	Twitter users	DT, NB, SVM-L, SVM-R	Twitter profiles were analyzed and reported higher accuracy with single ML classifier as SVM-L.	82
2019	Tadesse et al.	1,841	Reddit	SVM, MLP	In Reddit posts, utilize a model to find any elements that could indicate relevant online user's depressive views.	90
2021	Syms and Raj	2,500	Twitter	DT, RF, SVM, NB, SVM-NB(Hybrid)	Create hybrid model with two classifiers SVM and NB to get improved results.	92
2021	U M Haque et al.	6,310	YMM for Child mental health	RF	Reported 11 parameters to identify depression among children and adolescents and tested with individual classifier RF.	95
2021	Arachchige et al.	1,335	Facebook, Twitter and Reddit	ML and NLP	Study is reproted over three different online media platforms with the use of NLP and ML techniques.	-
2020	Mali et al.	13,321	Reddit	MNB, SVM, KNN	Creating a topic model to find topics that are hidden but serve as depression triggers.	89.36
2020	Zainab et al.	20,000	Reddit	LR, RF	Reddit text data on depression and non-depression were analyzed using ML and explainable in both Urdu and English.	86

2020	Trifan et al.	9,210	Reddit	SVM, PAC, MNB, SGD	A model based on specially created psycholinguistic traits as potential enhancements to current methods of classifying depressed online personas.	82
2018	Lynos et al.	463	Online forums (mentalhealthforum.net, psychoforums.net)	Statistical Analysis	Model for identifying specific verbal tics used by the people with mental illness and other issues online interaction and communication.	-
2021	Garg P. et al.	Chest & Wrist devices over 13 subjects	Wearable Stress and Affect Detection (WESAD) public dataset	KNN, LDA, RF, Adaboost and SVM	Stress detection using wearable sensors has been evaluated but study limited to imbalanced public dataset.	84.17
2023	Vasha et al.	10,000	Facebook and YouTube	SVM, LR, DT, RF, KNN and NB.	Suicidal ideation using six ML classifiers but achieved best results with individual classifier as SVM.	75.15

PAD: Passive Aggressive Classifier; SGD: Stochastic Gradient Descent; MNB: Multi-modal NB; YMM: Young Minds Matter

3. Materials and Methods

This section discusses the proposed methodology adopted for the identification of neurological disorder symptoms among the posted content over social media network. The various embedding and classification algorithms are discussed. The proposed architecture of the work is discussed below in Fig.1.

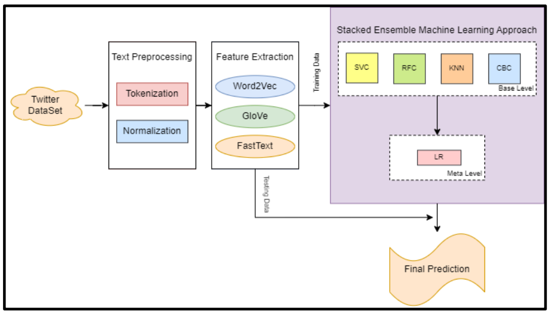


Fig. 1. Framework of the proposed work comprises Twitter dataset, text preprocessing, feature extraction, model and final prediction phases.

3.1. Dataset Preparation

Twitter dataset is chosen for this study and downloaded from Kaggle, an open-source platform for dataset collection [44]. Today, the most effective instrument for qualitative analysis of data is API. It makes it possible to organize the data, deconstruct, and extract knowledge from data. Any type of data can be used. It may take the form of social media posts or open-ended survey responses [39]. For the research to be accurate, a suitable dataset must be picked. The amount of the data, its accuracy, and integrity are some factors that are crucial to the research [40].We have taken the dataset of Twitter comments, 10308 comments. In this category we count for the total number of neurological disorder detection as label 0 and 1 respectively shown in Fig. 2.

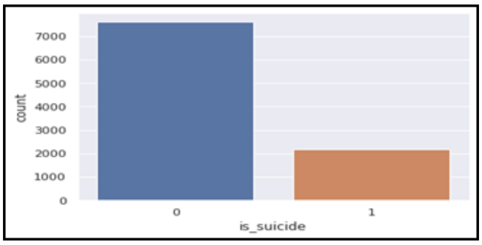


Fig. 2. Twitter dataset classification with two class labels.

Table 2 displays some of tweets before preprocessing as posted on Twitter.

Table 2. Text before Preprocessing displays message to examine and label results as 0 or 1.

tweets.head( )			
Index		**Message to examine**	**Label**
0	106	Just had a real good moment. I missssss hi..	0
1	217	is reading manga. http://plurk.com/p/mzp1e..	0

2	220	@comeagainjen. http://twitpic.com/2y2lx...	0
3	288	@lapcat Need to send ‘em to my accountant.	0
4	540	ADD ME ON MY SPACE!!! myspace.com/LookThunder..	0

tweets.tail( )

Index		Message to examine	Label
10309	802309	No depression by G Herbo is my mood from now ….	1
10310	802310	What do you do when depression succumbs…	1
10311	802311	Katamine Nasal Spray shows promise against …	1
10312	802312	Don’t mistake a bad day with depression !	1

3.2. Text Classification

For performing text classification, data needs to be pre-processed using tokenization and normalization methods that includes stemming and lemmatization to get cleaned text as shown in Table 3 and Table 4 respectively.

3.2.1. Preprocessing

As the text is least structured type of data, resulting in a large amount of cleaning. To extract the accurate information from the text, with the help of these preprocessing approaches, noise in high dimensional data features can be easily converted to low dimensional data features. Depending on the information and the situation, preprocessing data may involve a number of processes.

3.2.2. Tokenization

Text is first transformed into the tokens, and then into vectors is known as the process of tokenization. Words, characters, numerals, symbols, and n-grams are a few examples of tokens. Additionally, the removal of pointless tokens is made simpler. For instance, documents can be split into parts or sentences. We are word-tokenizing the reviews in this instance. Whitespace/unigram tokenization is the most popular tokenization technique. By separating the words from the whitespace, the entire text is divided into words in this procedure.

Table 3. Tokenized Text.

	Clean_message	Is_suicide	Tokenized_text
6069	secretsocietydi hope ur great don’t spend cash	0	[secretsocietydi, hope, ur, great, don’t, spend…
2154	allanzzz oh allan may god give strength energi…	0	[allanzzz, oh, allan, may, god, give, strength…
8431	slip anoth depress	1	[slip, anoth, depress…
9087	removsomebodiels depress anxieti help emoji…	1	[remov, somebodi, els, depress, anxieti, help…
4967	britneyfr regular tri find food noth new	0	[britneyfr, regular, tri, find, food, noth, new]

3.2.3. Normalization

Correct processing must be applied to words that appear different due to case or are written in a different way but possess same meaning. Processes of normalization guarantee that all these phrases are treated similarly. For instance, changing the case of all text or transforming numerals to their word equivalents. A text can be made clean by using normalization to reduce the amount of unique token, eliminate variances, and get rid of extraneous information. Lemmatization and Stemming are the two often used techniques for normalizing the text.

3.2.4. Stemming

The method for removing inflationary forms from a given token in a simple rule based one. The result of the error is a word’s stem. For instance, after stemming, the words like *laughing, laughed, laughs and laugh* all form *laugh.*

3.2.5. Lemmatization

In this work, the lemmatization method from the Gensim package, which handles lower case, numeric elimination, self-contained commas, special characters, and punctuation, was used.

When we consider a corpus W that consists of pre-processed words through stemming forms a list as [41].

W = {word1, word2, word3………,wordk}

This list is created in such a manner that the occurrence of every word is only one times, in such a manner (1).

$$word_m \neq word_n \; for \; m \neq n \qquad (1)$$

For the data that consists of processed text in the form of dictionary documents in (2),

$$D = \{D_1, D_2, \dots \dots \dots, D_p\} \qquad (2)$$

Hence for dictionary $\mathbb{D}$, we formed an induced corpus as in (3):

$$W = unique\left(\bigcup_{m=1}^{p} D_m\right) \tag{3}$$

where "unique" function implies that repeated terms are eli minated, resulting in a union of texts as a dictionary.

Table 4. Cleaned Text after Preprocessing.

		Message to examine	Is_suicide	Clean_message
0	106	just had a real good moment I missssssss hi..	0	real good moment missssss much
1	217	is reading manga http://plurk.com/p/mzp1	0	read manga
2	220	@comeagainjen http://twitpic.com/2y2lx	0	comeagainjen
3	288	@lapcat Need to send'em to my accountant tomo…	0	Lapcat need send em account tomorrow
4	540	ADD ME ON MY SPACE!!! myspace.com/LookThunder	0	add myspacmyspacecomlookthun

After text preprocessing, we have a collection of p documents and dictionary of k words. Creating a term-document matrix $X \in \mathbb{R}^{nxp}$ listed as in (4).

$$x_{ij} = \text{total word}_i \text{ occurence in } D_j \quad 1 \le i \le n, 1 \le j \le p \tag{4}$$

This is how all terms are stored in numerical vector form as matrix in (5),

$$X = \left[x^{(1)} x^{(2)} \ldots\ldots\ldots x^{(p)}\right] \quad x^j \in \mathbb{N}^n \tag{5}$$

Where, single document is represented by $x^{(j)}$.

3.3. Feature Extraction

The selection of features for machine learning model should only include those that are essential. The efficiency of the model may be harmed by choosing irrelevant features. The process of feature selection aids in the elimination of redundant and pointless characteristics that do not improve the model's overall performance [42], [43]. The primary premise is that related words should have near values in numerical vectors representing them if their contexts are similar. The king-queen relationship is frequently used to define this concept in (6).

$$v_{king} - v_{man} \approx v_{queen} - v_{woman} \tag{6}$$

where vf is the vector representing token t.

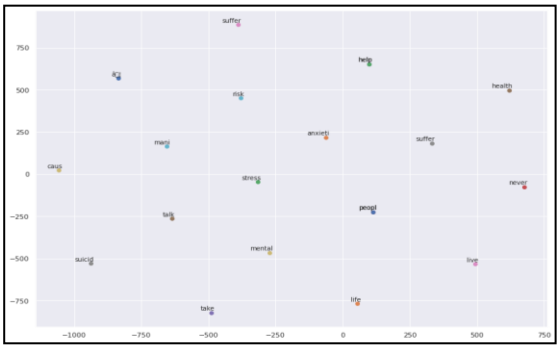
The unidentified keyword being searched for thus meets the following relation thanks to effective word embedding as (7) or (8) below.

$$v_{unknown} - v_{complex} \approx v_x - v_{real} \tag{7}$$

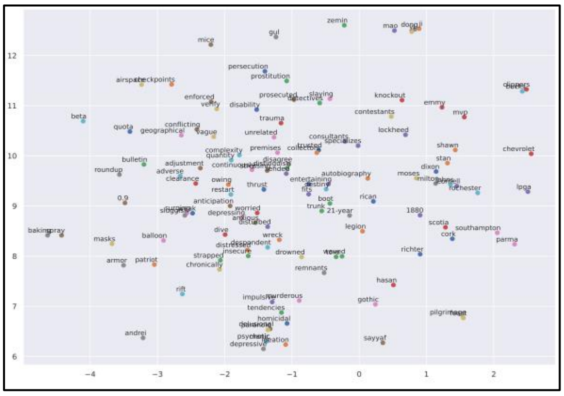
or similarly,

$$v_{unknown} = v_x + \left(v_{complex} - v_{real}\right) \tag{8}$$

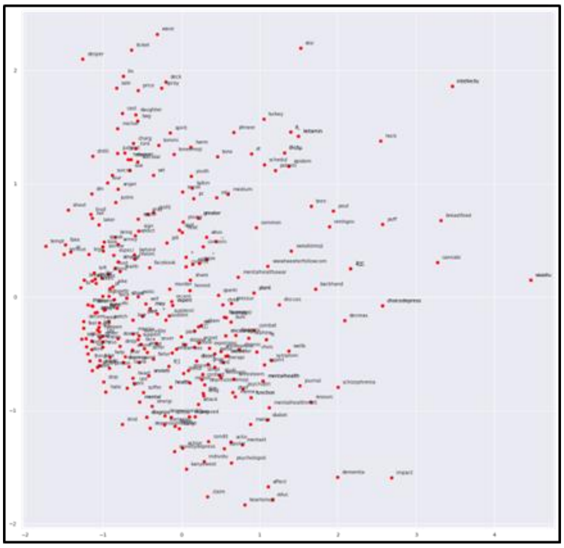
In order to find the unknown term, one must locate the vectors that are closest to the real vector and then get words that correspond to those vectors. The unidentified phrase should ideally be the top match or at least one of the below few matches with good embedding. Using word embeddings models as word2vec, GloVe and Fasttext vectors are created in this study that displays closed word found in the dataset as shown below in Fig. 3.



(a)



(b)



(c)

Fig. 3. Vector creation using three different word embedding techniques (a) Word2Vec (b) GloVe and (c) Fasttext.

4. Proposed Stacked Ensemble Algorithm

To enhance the generalization and predictions, ensemble learning integrates the results of individual models. A singular model trained on biased data is less likely to be used when using ensemble approaches, which lessen the chance of miss representation by merging numerous models [24]. Unlike ensemble methods, which may carry out random seed with varying start points with fewer computer resources, most training methods search locally for a solution which in turn confines the best answer [45-47].

The process of Ensemble Learning in mathematical terms looks like as (9):

$$\mathrm{Ensemble}(\mathrm{D}, \mathrm{M}, \mathrm{R}, \mathrm{A}) = S\left(\sum_{1}^{n} \mathrm{I}(\mathrm{D}, \mathrm{M}_{\mathrm{n}}, \mathrm{R}, \mathrm{A})\right) \quad (9)$$

where, n stands for number of models; S for Stacked model; I for Individual model; D stands for data; M for Models; R for Results and A for Analysis.

Basically, the stacked ensemble approach follows the basic steps are:

1) The dataset D = (xi,yi) where i = 1 to n is split into the training set Da & the test set Db. Here, xi represents the ith sample's feature vector, yi represents its predicted value, and n represents the dataset's sample size.

2) Using the training dataset that the first layer selects, the L model base learners train the model. L base students and trains the first layer model using cross validation CV. The dataset in (10) as;

$$\mathrm{C_s} = \mathrm{y_i}, \mathrm{p_{1,i}}, \mathrm{p_{2,i}}, \mathrm{p_{3,i}} \ldots\ldots \mathrm{p_{L,i}} \quad (10)$$

is obtained by combination with yi, where i=1,2,.....n/m.

The feature vector xi in the test set is used to determine the predicted value for each sample.

Let $$\mathrm{S} = \mathrm{z_1}, \mathrm{z_2} \ldots\ldots \mathrm{z_{N/M}} \quad (11)$$

Equation (3) is second layer meta-input learner data. The Meta learner is then fed with the K predicts results obtained from the experiment set training along with the average expected of such K prediction model in the test set.

3) To obtain the final prediction, Meta-learner predicts the first layer's input data. The base learner's prediction mistakes are corrected by the Meta-model, and to avoid over fitting, the whole set of data is only used once during the model training. The meta-training does not use any of the data items that the base classifier predicted.

In this extended work of [49] & [50], stacked ensemble of four base classifiers as Support Vector (SVC), Random Forest (RFC), K-Nearest Neighbor (KNN) and Catboost classifier (CBC) are used with a Meta classifier as Logistic Regression (LR). SVC, RFC, KNN, and CBC are just few of the four widely used predictors that are combined in this paper that proposed stacked ensemble learning approach.

The expected value can then be expressed as at time t in (12) as,

$$\hat{\mathrm{Y}}_{\mathrm{Ensemble(t)}} = \hat{\mathrm{Y}}_{\mathrm{SVC(t)}} + \hat{\mathrm{Y}}_{\mathrm{RFC(t)}} + \hat{\mathrm{Y}}_{\mathrm{KNN(t)}} + \hat{\mathrm{Y}}_{\mathrm{CBC(t)}} \quad (12)$$

Where, $\mathrm{t} = 1,2, \ldots \mathrm{m}$.

To improve the efficiency of our model, predictions from base classifiers are then stacked to create ensemble model [48]. It performs stratified folding with 10-fold CV which is then tuned with Meta classifier over some parameters to get improved accuracy as explained in *Algorithm 1* for ensemble modeling approach.

Algorithm 1

Step1: Select the dataset from Kaggle source as Sentiment Analysis for Tweets.

Step2: Divide the dataset into training and testing dataset.

Step3: Perform Text preprocessing and feature extraction algorithms to get cleaned text.

Step4: Divide training sets into n-folds by using RepeatedStratified10Fold. First fold, which would be n-1, has now been fitted to the base learners as SVC, RFC, KNN and CBC, and it will now generate prediction for the nth fold.

$$\widehat{Y}_{\mathrm{Ensemble(t)}} = \widehat{Y}_{\mathrm{SVC(t)}} + \widehat{Y}_{\mathrm{RFC(t)}} + \widehat{Y}_{\mathrm{KNN(t)}} + \widehat{Y}_{\mathrm{CBC(t)}}$$

Step5: The x1 train list is updated with the prediction given in the previous phase.

Step6: Steps 2 and 3 should be repeated for the remaining n-1folds to get an array of size n called x1 train.

Step7: The model has now been trained on all n parts and can now predict the results of the sample data.

Step8: Include prediction in the y1 test list.

Step9: By using Models two and three for training, respectively, we may determine x2 train, y2 test, x3 train, and y3 test to obtain Level 2 predictions.

Step10: Now, learn the Meta model using level 1 prediction. The model will use these predictions as features.

Step11: Finally, predictions on testing data in the stacked model may now be made using meta learner as Logistic Regression.

Step12: Classification of tweets as final output with improved accuracy scores.

5. Results and Discussion

On a selected dataset, the stack-based ensemble classification model as shown in Fig. 1 is used. This hybrid model based on stack-based ensemble of SVC, RFC, KNN and CBC as base classifier is created using LR Meta learners. The proposed model shows improved accuracy scores with hybrid stacked ensemble model over three-word embeddings as shown in Fig.4. The graphical representation of performance analysis bases on execution time is also shown in Fig. 5. In addition, the proposed study is superior to existing literature with improved accuracy scores can be shown in Table 5 and 6 below.

The proposed hybrid model comprising an ensemble of four base learners with LR Meta learners provide better sustainable improved accuracy of *0.984817* with word2Vec, *0.996134* with GloVe and *0.980570* with Fasttext word embedding models. The individual scores for all base models are well explained and comparison is then discussed in Table 5. When compared to individual machine learning methods, ensemble classifiers also outperformed them introducing stratified folding on stacked model to the feature set increases their performance for the sentiment analysis. Confusion matrix and RoC Curve for the proposed study are mentioned in Fig. 6 and 7 respectively. It states that proposed model calculates better results with GloVe embedding rather than other two opted in this study.

The performance time is also evaluated during the study. It is also noticed that execution time is higher for proposed stacked ensemble classifier model as compared to four single machine learning classifiers. Hybrid stacked model took more time to perform but producing improved results as compared to single machine learning approach. This study also calculates the metric evaluation scores as discussed in Table 7 in terms of precision, recall, and F1-scores.

Table 5. Comparison of accuracies for all embeddings of proposed model over standalone learning models.

Model Classifier	**Word2Vec**	**GloVe**	**FastText**
SVC	0.979709	0.992328	0.974957
RFC	0. 958296	0. 990606	0. 961481
KNN	0. 973559	0. 963113	0. 94788

CBC	0. 975322	0. 995153	0. 979749
Proposed Stacked Model	**0. 984817**	**0. 996134**	**0. 98057**

Table 6. Performance Comparison of the proposed ensemble model with the existing literature cited in this study.

Research	Total Users	Dataset	Approach	Accuracy (%)
Na et al. (2020)	6588	Korea Users	Machine Learning	86.20
Sau and Bhakta(2017)	520	Geriatric patient	Random Forest	89
Hatton et al. (2019)	284	Geriatric patient	XGBoost	74
Jiménez- Serrano et al. (2015)	1397	Postpartum period mums	Naive Bayes	79
Wang and Li (2019)	400	Mammograms breast images	CNN and ELM	86.5
Sri (2020)	303	Cleveland heart	Voting Ensemble	90
Jaymin Patel(2017)	303	Cleveland heart	J48, RF and LR	56.76
RinkiChhaterjee (2021)	7146	Facebook	Naive Bayes	76.6
Alsagri and Ykhlef (2020)	500	Twitter	SVM-L	82
Proposed Stacked Ensemble (2024)	**10308**	**Twitter**	**Ensemble Learning**	**99.6**

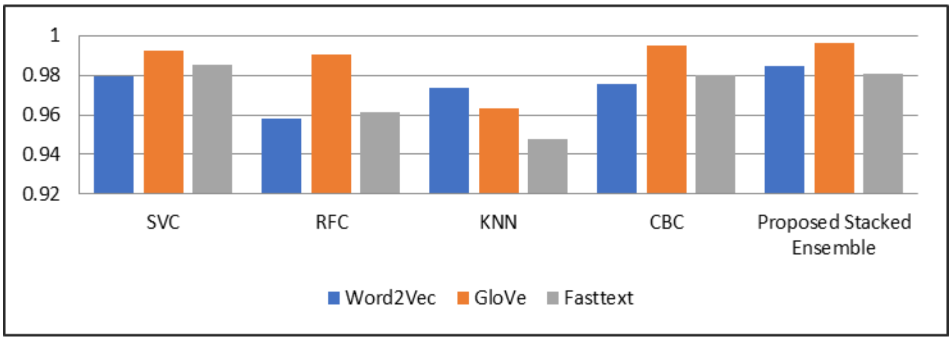


Fig. 4. Graphical representation of accuracy scores of proposed work.

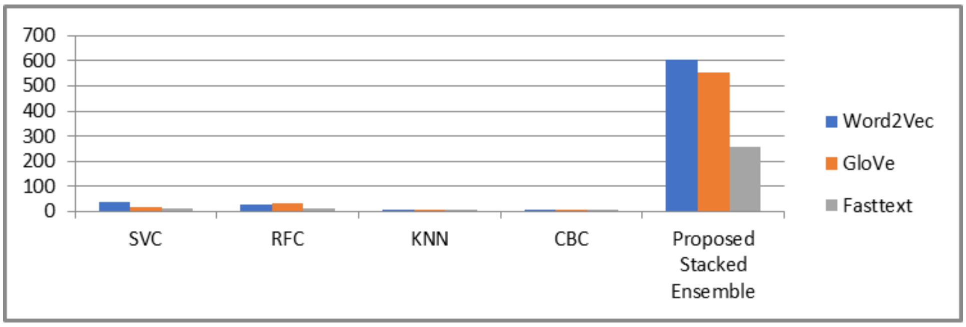
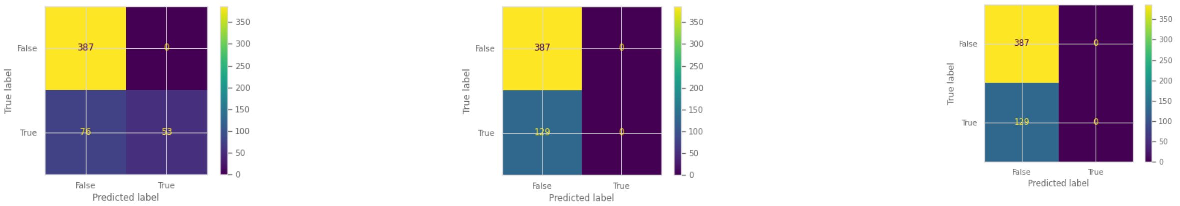


Fig. 5. Comparison based on the execution time for standalone models and stacked ensemble model.

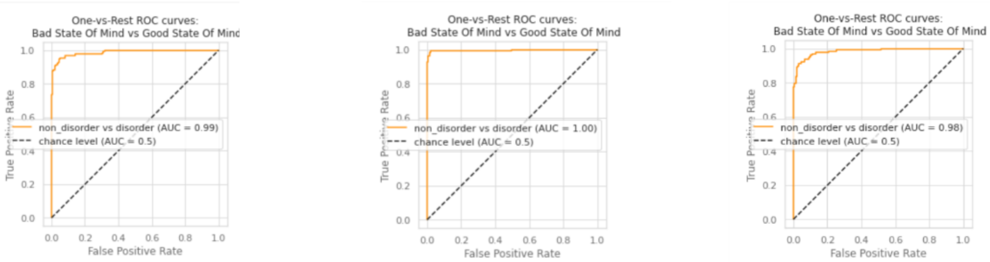
Table 7. Metric Calculation for proposed model in terms of Precision, Recall and F1 score for both the classes.

		Precision	Recall	F1 Score
Word 2Vec	Class 0	0.835	1.0	0.9105
	Class 1	1.0	0.4108	0.5824
GloVe	Class 0	0.75	1.0	0.8572
	Class 1	0.0	0.0	0.0
Fasttext	Class 0	0.75	1.0	0.8571
	Class 1	0.0	0.0	0.0



(a) (b) (c)

Fig. 6. Confusion Matrix for the proposed model with (a) Word2vec, (b) GloVe and (c) Fasttext embeddings.



(a) (b) (c)

Fig. 7. RoC Curve of the proposed stacked ensemble model with (a) Word2Vec, (b) GloVe and (c) Fasttext embedding methods.

Hence, when we apply sentiment analysis with this proposed approach for classification of posted content on such social media platform, this proposed ensemble model correctly classified tweets as positive or negative activity on the web that helps in assisting any type of neurological disorder.

6. Conclusion and Future Work

Depression, a neurological disorder can develop in a person for a variety of reasons. First, a dataset of 10308 participants has been conducted to screen for early identification of neurological disorder over social networking profiles that affects human psychological resilience. The aim of this research was to identify neurological disorder as depression over tweets using hybrid stacked ensemble learning classifiers. The stacked classifier with the stratified folding cross validation strategy is better model to predict disorder among the social identity, as shown by the results of the chosen algorithms considered in this study. It is clearly noted that proposed stacked ensemble model achieved higher accuracy over all single models for all the chosen word embeddings; Word2Vec, GloVe and Fasttext as 98.4%, 99.6% and 98.05% respectively. Hence, our proposed stacked ensemble model outperforms over traditional ML models.

This work can be extended to various embedding models over other classifiers with huge dataset as compared to chosen in this study. Additionally, the proposed model can be trained using image sentiment classification as a future direction.

Acknowledgements

We would like to give our sincere gratitude to the reviewers for their constructive comments.

Author contributions

Tejaswita Garg: Conceptualization, Methodology, Software, Field study, Data curation, Writing-Original Draft preparation, Validation, Field Study.

Sanjay K. Gupta: Visualization, Investigation, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] Garg, M., 2023. Mental Health Analysis in Social Media Posts: A Survey. Archives of Computational Methods in Engineering, pp.1-24.

[2] M.M. Islam, M.R. Haque and M.N. Kabir, 2020. Breast Cancer Prediction: A Comparative Study using Machine Learning Techniques. SN Computer Science, Vol. 1, No. 5, pp. 1-14.

[3] WHO, 2021. Depression and Other Common Mental Disorders: Global Health Estimates. Available online at: https://apps.who.int/iris/handle/10665/254610.

[4] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., et al., 2017. "Depression detection via harvesting social media: a multimodal dictionary learning solution" in Proceedings of the Twenty-Sixth *International Joint Conference on Artificial Intelligence (Melbourne, FL: IJCAI-17)*, 3838–3844.

[5] Whitley, R., and Denise Campbell, R., 2014. Stigma, agency and recovery amongst people with severe mental illness, Soc. Sci. Med. 107, 1–8. doi: 10.1016/j.socscimed.2014.02.010.

[6] World Health Organization, 2021. Eastern Mediterranean Region.

[7] Whooley, M.A. , Wong, J.M. , 2013. Depression and cardio vascular disorders. Annu. Rev. Clin. Psychol. 9, 327–354.

[8] Otte, C. , Gold, S.M. , Penninx, B.W. , Pariante, C.M. , Etkin, A. , Fava, M. , Mohr, D.C. , Schatzberg, A.F. , 2016. Major depressive disorder. Nat. Rev. Dis. Primers 2 (1), 1–20 .

[9] Wang Y, Bao S and Chen Y, 2023. How does social media use influence the mental health of pancreatic cancer patients: a chain mediating effect of online social support and psychological resilience. Front. Public Health 11:1166776. doi: 10.3389/fpubh.2023.1166776.

[10] Orrù, G. , Monaro, M. , Conversano, C. , Gemignani, A. , Sartori, G. , 2020. Machine learning in psychometrics and psychological research. Front. Psychol. 10, 2970.

[11] Byeon, H., 2021. Predicting the severity of Parkinson's disease dementia by assessing the neuropsychiatric symptoms with an SVM regression model. Int. J. Environ. Res. Public Health, 18, 2551.

[12] Byeon, H., 2021. Comparing ensemble-based machine learning classifiers developed for distinguishing hypokinetic dysarthria from presbyphonia. Appl. Sci., 11, 2235.

[13] Kandel I, Castelli M, Popoviˇc A, 2021. Comparing stacking ensemble techniques to improve musculoskeletal fracture image classification, J Imaging. 7:100. doi: 10.3390/jimaging7060100.

[14] Byeon, H., 2021. Exploring factors associated with the social discrimination experience of children from multicultural families in South Korea by using stacking with non-linear algorithm, Int. J. Adv. Comput. Sci. Appl, 12, 125–130.

[15] Byeon H., 2021. Exploring factors for predicting anxiety disorders of the elderly living alone in South Korea using interpretable machine learning: a population-based study. Int J Environ Res Public Health 18:7625. doi: 10.3390/ijerph18147625.

[16] Jayawickreme N, Atefi E, Jayawickreme E, Qin J, Gandomi AH., 2020. Association rule learning is an easy and efficient method for identifying profiles of traumas and stressors that predict psychopathology in disaster survivors: the example of Sri Lanka. Int J Environ Res Public Health, 17:2850. doi: 10.3390/ijerph17082850.

[17] Na, K.S. , Cho, S.E. , Geem, Z.W. , Kim, Y.K. , 2020. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. Neurosci. Lett. 721, 134804.

[18] Sau, A. , Bhakta, I. , 2017. Predicting anxiety and depression in elderly patients using ma- chine learning technology. Healthc. Technol. Lett. 4 (6), 238–243.

[19] Hatton, C.M., Paton, L.W. , McMillan, D. , Cussens, J. , Gilbody, S. , Tiffin, P.A. , 2019. Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare, J. Affect. Disord. 246, 857–860 .

[20] Jiménez-Serrano, S., Tortajada, S., García-Gómez, J.M., 2015. A mobile health application to predict

postpartum depression based on machine learning. Telemed. e-Health 21 (7), 567–574 .

[21] Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H, et al., 2019. Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features, IEEE Access;7:105146–58.

[22] Sri, B.U., 2020. Effective Heart Disease Prediction Model Through Voting Technique. *International Journal of Engineering Technology and Management Sciences (IJETMS),* Issue: 5, 4, 10-13, September.

[23] Chatterjee R, Gupta RK, Gupta B., 2021. Depression detection from social media posts using multinomial Naive theorem, IOP Conf Ser Mater Sci Eng.; 1022:012095.

[24] Pirina I, Çöltekin Ç, 2018. Identifying depression on reddit: the effect of training data. In: Proceedings of the 2018 EMNLP workshop SMM4H: the 3rd social media mining for health applications workshop & shared task, Association for Computational Linguistics. https:// doi. org/ 10.18653/ v1/ w18-5903.

[25] H. S. Alsagri and M. Ykhlef, 2020. Machine learning-based approach for depression detection in twitter using content and activity features, IEICE Transactions on Information and Systems, vol. 103, no. 8, pp. 1825–1832, doi: 10.1587/transinf.2020EDP7023.

[26] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, 2019. Detection of depression-related posts in reddit social media forum, IEEE Access, vol. 7, pp. 44883–44893, doi: 10.1109/ACCESS.2019.2909180.

[27] S. Smys and J. S. Raj, 2021. Analysis of deep learning techniques for early detection of depression on social media network-a comparative study, Journal of Trends in Computer Science and Smart Technology, vol. 3, no. 1, pp. 24–39, doi: 10.36548/jtcsst.2021.1.003.

[28] U. M. Haque, E. Kabir, and R. Khanam, 2021. Detection of child depression using machine learning methods, PLOS ONE, vol. 16, no. 12, pp. 1–13, doi: 10.1371/journal.pone.0261131.

[29] I. A. N. Arachchige, P. Sandanapitchai, and R. Weerasinghe, 2021. Investigating machine learning & natural language processing techniques applied for predicting depression disorder from online support forums: A systematic literature review. Information, vol. 12, no. 11, pp. 1–18, doi: 10.3390/info12110444.

[30] Mali, A.; Sedamkar, R.R., 2020. "Prediction of depression using Machine Learning and NLP approach" *in Proceedings of the e-Conference on Data Science and Intelligent Computing*, Mumbai, India, 27–28 November 2020; pp. 46–50

[31] Zainab, R., 2020. Detecting and Explaining Depression in Social Media Text with Machine Learning. GOOD Workshop KDD'20, 1–4.

[32] Trifan, A.; Antunes, R.; Matos, S.; Oliveira, J.L., 2020. Understanding depression from psycholinguistic patterns in social media texts. In Advances in Information Retrieval; Springer: Cham, Switzerland; pp. 402–409.

[33] Lyons, M.; Aksayli, N.D.; Brewer, G., 2018. Mental distress and language use: Linguistic analysis of discussion forum posts. Comput. Hum. Behav., 87, 207–211. [CrossRef]

[34] Garg P, Santhosh J, Dengel A, Ishimaru S., 2021. Stress detection by machine learning and wearable sensors. International Conference on Intelligent User Interfaces, Proceedings IUI; pp. 43–45.

[35] Squires, M., Tao, X., Elangovan, S., Gururajan, R., Zhou, X., Acharya, U.R. and Li, Y., 2023. Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment. Brain Informatics, 10(1), pp.1-19.

[36] Vasha, Z.N., Sharma, B., Esha, I.J., Al Nahian, J. and Polin, J.A., 2023. Depression detection in social media comments data using machine learning algorithms. Bulletin of Electrical Engineering and Informatics, 12(2), pp.987-996.

[37] Zhang, T., Yang, K., Ji, S. and Ananiadou, S., 2023. Emotion fusion for mental illness detection from social media: A survey. Information Fusion, 92, pp.231-246.

[38] Herdiansyah, H., Roestam, R., Kuhon, R. and Santoso, A.S., 2023. Their post tell the truth: Detecting social media users mental health issues with sentiment analysis. Procedia Computer Science, 216, pp.691-697.

[39] Hung, L.P. and Alias, S., 2023. Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection. Journal of Advanced Computational Intelligence and Intelligent Informatics, 27(1), pp.84-95.

[40] Haralabopoulos, G., Anagnostopoulos, I. and McAuley, D., 2020. Ensemble deep learning for multilabel binary classification of user-generated content. Algorithms, 13(4), p.83.

[41] Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Proceedings of the 26th international conference on neural

information processing systems (NIPS), Volume 2 (pp. 3111–3119). USA: Curran Associates Inc.

[42] Febriansyah, M.R., Yunanda, R. and Suhartono, D., 2023. Stress detection system for social media users. Procedia Computer Science, 216, pp.672-681.

[43] Ghosh, T., Al Banna, M.H., Al Nahian, M.J., Uddin, M.N., Kaiser, M.S. and Mahmud, M., 2023. An attention-based hybrid architecture with explainability for depressive social media text detection in Bangla. Expert Systems with Applications, 213, p.119007.

[44] Shinigami, Dataset from kaggle repository (2021), https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets.

[45] Dorneles, S. O., Francisco, R., Barbosa, D. N. F., & Barbosa, J. L. V., 2022. Context Awareness in Recognition of Affective States: A Systematic Mapping of the Literature. International Journal of Human–Computer Interaction, 39(8), 1563–1581. https://doi.org/10.1080/10447318.2022.2062549.

[46] Zhou, Y., Kumar, A., Gandhi, C. P., Vashishtha, G., Tang, H., Kundu, P., Singh, M., & Xiang, J., 2023. Discrete entropy-based health indicator and LSTM for the forecasting of bearing health. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 45(2). https://doi.org/10.1007/s40430-023-04042-y.

[47] Ganapathi, I. I., Javed, S., Ali, S. S., Mahmood, A., Vu, N. S., & Werghi, N., 2022. Learning to localize image forgery using end-to-end attention network. Neurocomputing, 512, pp 25–39. https://doi.org/10.1016/j.neucom.2022.09.060.

[48] Tejaswita Garg, Sanjay K. Gupta., 2023. A Novel Algorithm for Stacked Generalization Approach to Predict Neurological Disorder over Digital Footprints. *International Journal of Modern Education and Computer Science (IJMECS)*, Vol.15, No.5, pp. 60-73, DOI:10.5815/ijmecs.2023.05.05.

[49] Garg, T. and Gupta, S.K., 2022, April. Efficient approaches to predict neurological disorder using social networking sites. In 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT) (pp. 294-298). IEEE.

[50] Garg, T. and Gupta, S.K., 2022, December. A Hybrid Stacked Ensemble Technique to Improve Classification Accuracy for Neurological Disorder Detection on Reddit posts. In 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN) (pp. 256-260). IEEE.