



Development of intelligent system based on synthesis of affective signals and deep neural networks to foster mental health of the Indian virtual community

Mandeep Kaur Arora¹ · Jaspreet Singh² · Anju Singh³

Received: 26 July 2022 / Revised: 13 November 2023 / Accepted: 2 December 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2024

Abstract

Social network analytics in health-care is recently catching-all as world is recovering from Covid-19 pandemic and when it comes to dealing with mounting mental health problems in the online community. A technical breakthrough to quickly address this issue is a big bet on researchers to deliver better intelligent mental health scale and support in mental distress risk management. In this paper we aim to build four artificial-intelligence driven models by blending the power of two dominant deep learning neural networks for explaining and predicting mental distress risk in the crowd-sourced online community. The models are simulated using a novel feature construct, which is a combination of numerical and textual data. The numerical data are realized by encoding social networking sites behavior and physical, social, cognitive experiences as part of three axes of psychological distress (depression, anxiety and stress). The textual part of data is made up of social network comments mined to acquire mental distressed cues by applying feature extraction techniques such as word embeddings and glove embeddings techniques. With the hyper-parameter tuning of models, we attained excellent performance (accuracy ~ 99.20%) which proves the efficacy of the proposed hybrid mental distress prediction model well with respect to accuracy in comparison to other related recent existing models with a boost of 0.20%. Our experimental results offer a robust model wherein we achieved zero false positive cases, attained 100% precision and excellent F1 measures. Additionally, we validated our results by using state-of-the-art technique BERT on different ground truth dataset i.e., Indian tweets to explore the tweets for psychological-distress prediction. Thus, we present an effectual automated tool for treatment activation of mental distress and supporting decisions in fostering the mental health of online society.

Keywords Mental distress · Deep learning · Glove embeddings · Neural network · Machine learning

1 Introduction

Mental health is the most ignored area of health globally due to the poor consciousness of mental health signs, stigma associated with it, outdated system of services available, lack of preventive measures and shortage of human resources.

This was the fact before Covid-19, but the pandemic has further worsened the mental health stats worldwide. Mental health numbers are staggering, in a recent WHO update close to 1 billion people are living with mental disorders with India reported for nearly 15% of the global mental, neurological and substance abuse disorder burden¹. The access to quality mental health services health is low relatively to other health services. In low and middle-income countries, more than 75% of people with mental, neurological and substance use disorders do not receive treatment for their condition at all.¹ In India, the mental illness interference gap is almost 70 to 92% depending on the state, even when in a meta-analysis of community survey it is projected that

✉ Jaspreet Singh
jsarora1@gmail.com

Mandeep Kaur Arora
Mkaur.Arora1@gmail.com

Anju Singh
anjusingh40@gmail.com

¹ Kamla Nehru College, Delhi University, New Delhi, India

² Sharda University, Greater Noida, Uttar Pradesh, India

³ GD Goenka University, Gurugram, Haryana, India

¹ <https://www.who.int/news/item/27-08-2020-world-mental-health-day-an-opportunity-to-kick-start-a-massive-scale-up-in-investment-in-mental-health#:~:text=Mental%20health%20is%20one%20of,every%2040%20seconds%20by%20suicide>

the pervasiveness of depression and anxiety could be up to 33 per 1,000 persons as reported by WHO (*Mental Health in Asia*). India congregates 17.7% of the global population but very few survey studies have contributed to assessment of the incidence of mental disorders in India (Sagar et al. 2017).

Social Networking sites (SNS) have become a part of many people's daily lives as the popularity of virtual connectivity and messaging apps has skyrocketed. The increasing popularity of SNS services has resulted in harmful usage and subsequently opens a way to increase in psychological mental problems in online community. Literature talks greatly about the connection of usage of social media platforms with mental health threat (Karim et al. 2020). Previous research also found that mental health issues due to social media is prevalent in all ages but with respect to gender, females were much more likely to experience mental health than males (Iannotti et al. 2009) (Sagar et al. 2017).

One of the best ways to solve this problem is to investigate the individual's activities in greater depth. It is a challenge as well as responsibility of researchers and health professionals to find efficient ways to detect the mental health risk as early as possible to foster the mental health of the online community. In (Jo et al. 2017) authors highlighted the link between the computational models and psychological state, suggesting computational framework as fine objective method of assessing the real emotions of the person.

Since a decade the literature argues that Artificial Intelligence(AI) approach have advantages and are more beneficial with report of good performance in diagnosis of mental illness over traditional approaches. In Table 1, we present the comparison on points accentuated in the literature describing both the approaches, we emphasize that it can be considered as key motivation for the interested researchers in the field of fostering the mental health of the online community (Shatte et al. 2019) (Su 2020) (Saqib et al. 2021) (Thieme et al. 2020) (Cho et al. 2019) (Chancellor and De Choudhury 2020).

This paper establishes a framework for evaluating psychological distress episodes in the online society by employing a novel feature construct consisting of person's overall personal, social and cognitive behavior on social media. The primary aim of this experiment is to put forward a reliable and unflinching processing model that can find the most significant mental risk features from the crowd-sourced dataset to improve the prediction accuracy and lessen the burden on clinicians and healthcare professionals. Following are the main contributions of the paper:

- The proposed work assimilates a crowd-sourced data particularly from Indian online community with a novel set of feature construct consisting of some essential general information, social networking sites behavior, physi-

cal, social and cognitive experiences gathered through three axes of psychological distress (depression, anxiety, stress) combined with logs/postings on social media. We propose that our original dataset can be considered as benchmark for further research in this area.

- The capabilities and strengths of two leading neural network techniques namely CNN and LSTM are hybridized to develop an automated system to predict mental distress happening in the virtual community and to identify the influential risk factors for mental distress that contribute the most to the diagnosis and treatment.
- Our work utilizes word embeddings and Glove embeddings techniques, a dense vector representation of words that captures some context of the words on their own to process the logs/texts posted by the online society members and predict their mental distress state. Thus it contributes toward the importance of natural language processing tools in leading to cues related to mental health risk.
- The study contributes to depict the relationship between various chosen variables with the mental state and explains the correlation between significant factors as causal for psychological distress occurrence. The findings of the dataset examined in the proposed approach is in sync with what literature asserts, thus strengthens and validates the crowd-sourced dataset used for investigating the occurrence of mental distress
- Lastly, we offer an effective decision-making tool to assist clinicians and health caretakers in the early diagnosis of mental distress risks prevalent in the online society today.

The rest of the paper is aligned as follows: continuation to this section we present the problem statement followed by motivation for this study. Section 2 contains the literature review and related work connecting to this study. Sections 3 describe the methodology consisting of data pre-processing, methods, techniques and proposed algorithm. Section 4 presents results and discussion. Finally conclusion is discussed with future scope and mentioned in Sect. 5.

1.1 Problem Statement

To address the problem of deteriorating mental health of the Indian virtual community due to imprudent online behavior is crucial in the present scenario. This study develops robust Machine Learning models for accurate detection of mental disorders thereby attempting to preserve the mental health of the virtual community. The problem statement for the proposed study is mentioned as follows:

“Mental health issues are part and parcel of imprudent online behavior and may pose a serious threat to vir-

Table 1 A comparison of Traditional approach with AI approach of mental health detection in virtual community to drive adoption of automated systems based on AI and mining of Social Network data in the mental health ecosystem

Comparison points	Traditional approach	AI approach
Data Source	Questionnaire, Health records, clinicians or health practitioner diagnostic tool records etc	Questionnaire, self-report, crawl data from virtual platforms, electronic media, application interface tools etc
Information gathered	Tedious and prone to manipulations. Information gathered is incomplete as individuals are apprehensive and prefer not to disclose their true online behavior	Easy and noninvasively, automated process enables individuals to open up which results in honest sharing of feelings by individuals
Time and cost effectiveness	Time consuming activity as it requires clinician time and meantime the patient's state may worsen	Data collection is mostly inexpensive and does not require direct engagement of individuals. Automatic identification is faster and is a better way to deal
Intervention	The underlying social stigma hampers the diagnosis of mental illness results in delayed treatment	ML methods and techniques boost the development of reliable and effective diagnosis following early treatment
Intervention of experts	Requires domain expert for the result	Computer based learning enhances the work of experts and may result in novel outcome to manage the problem
Detection of relationship between independent and dependent variables	Manually interpreted	Automatic detection can identify interactions among the variables
Biasness and Analysis	Subject to bias due to the nature of the process and tiresome analysis impedes decision making	A step toward accurate, reliable and interpretable risk assessment tool supporting decision making
Missing data	Missing values cause difficulty in analysis and interpretation	Handles missing values efficiently and can be interpreted to match the true population
Outreach	Not possible to reach widespread population as difficult to access	Possible for widespread outreach and help in assessing mental health state through online interventions
Assessment of new case	The process need to be performed from the beginning for any fresh case to be identified for mental health risk	Learns the pattern from the data facilitates the assessment and classification of new cases at risk

tual society. The need of the hour is to develop a robust user-friendly inexpensive automated model for detection of mental disorders prevalent in online society”

1.2 Motivation

Overwhelming online community face various mental health problems like depression, anxiety, stress, loneliness etc. but they do not disclose their distress and avoid seeking help due to the social stigma or taboo attached to its incidence. The traditional methods of diagnosis of mental health of online society have inherent problems as discussed in Table 1. There is a need for reliable, efficient, in-expensive and non-intrusively automated system to preserve the well-being of the online society. Above mentioned tribulations motivated us to pursue this empirical study in order to play our part in contribution toward the upkeep of the mental health of online community.

2 Literature review

The affliction of mental health in online community has dominantly been discussed at various platforms by stakeholders and health caretakers. With the continued rise in popularity of online platforms in people’s lives in recent years, the research area involving three sectors—social media, health and big data analytics have come together and begun to evolve.

We divided the literature review in two parts: first part focused on general review of studies with the objective of answering the literature research questions (LRQ) mentioned as follows. The answers for these questions are derived from the review papers extensively focused on: machine learning in mental health (Shatte et al. 2019) (Thieme et al. 2020) (Rahman et al. 2020) (Kim et al. 2021), mental health monitoring using machine learning (Belfin et al. 2020) (Garcia-Ceja et al. 2018), predicting mental health from social media (Chancellor & De Choudhury 2020), application of machine learning methods in mental health detection (Abd Rahman et al. 2018), machine learning and natural language processing in mental health (Glaz et al. 2021), AI in mental health (D’Alfonso 2020).

LRQ1: What are the most common characteristics of the mental health that have been used for prediction?

We found following characteristics of the mental health used in various studies:

- (i) **Types:** suicide, psychosis, stress, depression, anxiety, obsessive compulsive disorder, post-traumatic stress disorder, bipolar disorder, eating disorder, anorexia, bulimia, schizophrenia, borderline personality dis-

order, internet addiction, cyber-relationships, cyber-bullying, information overload.

- (ii) **Thematic areas or domain for assembling data for prediction of mental health issues:** Traditional (Questionnaires), molecular, genomic genetic data, clinical, medical imaging, physiological signal to facial, body expressive and online Behavioral, Integration of multimodal data etc.
- (iii) **Social Platform:** Twitter, Facebook, reddit, live journals, Weibo, web forums, crowd-sourcing, Amazon Mechanical Turk, micro blogs.

LRQ2: What outcomes have been predicted in contributions related to mental health of SNS users?

The studies contributed toward the following outcomes: Risk levels of mostly depression, users’ behavior, attitude, pattern recognition by classifying users at risk/ no risk, brain and facial signals mapping with mental health etc. The literature lacks outcomes related to causal-effect and associations between factors pertinent in contributing toward the mental health risk.

LRQ3: What are the prediction features that are used to build prediction models in mental health detection?

Broadly we came across five types of data being discussed in the literature: (i) Demographic variables (ii) Biological variables: Related to family history, genetic and genomic profiles (iii) Clinical/Electronic health records (iv) Vocal and visual expression data (v) Social media data- Behavioral and linguistic cues from social media data.

LRQ4: What are the techniques/models used for prediction of mental health of SNS users.

In Literature classical machine learning and artificial neural network algorithms such as SVM, LR, RF, NB, Ada-Boost, DT, JRip Rule, Markov logic networks, Multinomial LR & NB, SVM with Radial Basis Function, Transductive SVM, MLP, CNN, LSTM etc. are applied for classification task. Recently researchers are building hybrid models to blend the capabilities of individual algorithms for better outcomes. This study utilizes blend of capabilities of CNN and LSTM and develops an efficient mental disorder detection system.

LRQ5: What metrics have been used to evaluate prediction results in mental health of SNS users?

To establish the usefulness of the studies in real situations researchers have mostly employed following strategies: (i) K-fold cross validation & hold-out method (ii) Machine learning metrics—accuracy, precision and recall, F1, sensitivity, specificity and area under the curve (AUC) (iii) few studies used regression-oriented measures as root mean squared error and *R* value.

In the second part of literature review we explored the recent works related to our study and found that there is high diversity in the machine learning algorithms chosen

by researchers. The most frequently used predictive algorithms were SVM, followed by logistic regression, random forest, decision tree, Naive Bayes, XGboost etc. Like in a paper (Haque et al. 2021) authors explored the performance of ML algorithms in the detection of depression in children. The researchers employed RF, XGBoost, DT and NB on a surveyed data and found that RF outperformed the other ML methods in classifying the depressed children class with 95% accuracy and 99% precision.

Currently, deep learning algorithms like CNN, RNN, and LSTM are trendier since they paved a way of better prediction accuracy, learn and get adapted to past experiences. This shift can be attributed to the characteristic of the deep learning algorithms to automatically grasp valuable features without relying on feature engineering and extraction. Literature also suggests that there has been an upward trend in the natural language processing techniques applied to mental illness detection (Zhang et al. 2022).

CNN based models have a general formation composed of a convolutional layer, pooling layer followed by a fully connected layer and are best known for automatic detection of significant features and efficient computational powers (L. Chen et al. 2020a, b). This characteristic is evident from the following studies related to our approach: Researchers in a study employed CNN on a large dataset from social platform to propose a factor graph combined with CNN using stress related texts, visuals, social attributes and achieved ~91% accuracy with 93.40 F1 measure (Lin et al. 2017). Another group of researchers used CNN approach to automatically recognize mental illness posts from Reddit and yielded 91.08% accuracy (Gkotsis et al. 2017). In one more study depression was identified in twitter users by using CNN architecture with optimized embeddings and the study reported accuracy, F1, precision and recall as 87.95%, 86.96%, 87.43% and 87.02% respectively (Husseini Orabi et al. 2018). Researchers in a study collected data from mental health related subreddits applied word embeddings and achieved best accuracy with CNN classification algorithm (Kim et al. 2020).

RNN-based structure is useful for text data as it retains the sequence of words and allows previous data to become inputs but has an inherent restriction of remembering long term dependency. However the extension of RNN such as LSTM and GRU neural network models can deal with it by memorizing the important information thus resolving the long-term dependency and returns better results. Investigators in a study used self-reported reddit posts for detection of depression and reported F score as 0.64 for LSTM model (Cong et al. 2018). The researchers (Ahmad et al. 2021) in a study deployed LSTM for analysis of sentiments (positive/negative) from Bangla text and achieved 94% accuracy. Another LSTM model study showed performance as

AUC = 0.94 for prediction of suicide using social media posts (Coppersmith et al. 2018).

For the task of detection of mental health issues, fusion of deep learning algorithms with NLP based feature extraction framework has shown better performances and attracted attention of researchers to explore and develop better ways to safeguard the virtual community from the future episode of mental health issues. General deep learning-based structures mainly enclose two layers: an embedding layer and classification layer. Embedding layer is obtained from the conversion of text to real-valued dense vectors which can keep sequence, connotation and context-based information. There are various embedding techniques such as ELMo, GloVe word embedding, word2vec, bidirectional encoder representations and ALBERT, used for analysis of the textual data by researchers (Ramírez-Cifuentes et al. 2021) (Syaputra & Ali 2022). The second layer i.e., classification layer utilizes various kind of deep learning algorithms like CNN, RNN and hybrid architecture to categorize the data into classes.

Recent craze for hybrid-based methods in literature combines several neural networks for mental illness detection such as CNN, RNN, encoder and LSTM, obtains local features along with long dependency features through NLP techniques and can outperform when they are used individually. In a recent study authors trained deep learning models by converting online posts to word embeddings and demonstrated that CNN + LSTM attained 97.65% accuracy and LSTM achieved 97.45% accuracy (Kanaan et al. 2021). Researchers in a study exploited twitter data for depression using a hybrid model (CNN-biLSTM) with word embeddings as input and obtained 94.28% accuracy (Kour & Gupta 2022). In one more recent study, word2vec and BERT with Bi-LSTM was examined by researchers to efficiently sense depression and anxiety indications from social media posts with an accuracy of 98% using the knowledge distillation technique (Zeberga et al. 2022).

In literature few recent studies also talked about potential of multiple features in identifying the multimodal mental disorders in SNS users. Like, the authors in a recent study demonstrated that the effective multimodal representation of features like audio-visual fused with textual data has capabilities to detect different mental disorders (Zhang et al. 2020). In another recent study (Ghosh et al. 2023) two publicly available MRI datasets (OASIS and ADNI) were utilized with low-cost transfer learning architecture (MobileNet) to detect Alzheimer's disease and researchers achieved 95.24% accuracy with OASIS dataset, 81.94% accuracy in ADNI dataset and 83.97% accuracy in both datasets(merged). In one new approach, authors (Ahmed et al. 2023) proposed a multimodal Neural network to detect multimodal depression by taking multiple features like video, audio, text and EEG. The incorporation of selective drop-outs, attention mechanism and normalization techniques resulted in greater

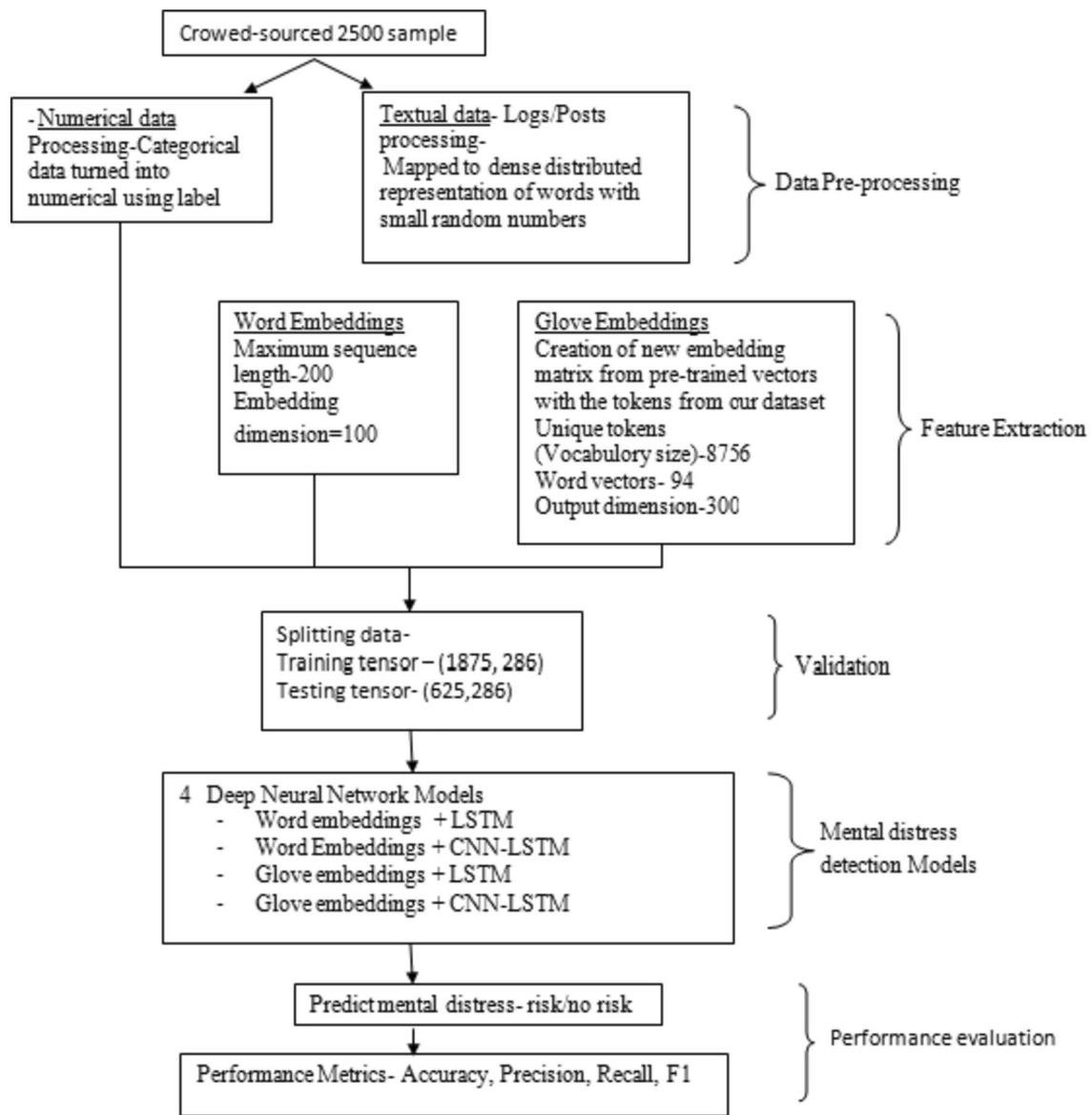


Fig. 1 The Phases of the proposed Approach

performance with accuracy as 0.96, F1-score as 0.94 and S1 score as 0.94.

According to the findings from the literature we infer that promising results have been achieved by the studies; however, several challenges exist for the automated task of mental illness detection which needs further research: obtaining annotated data for training purpose of the models is labor-intensive, time consuming and expensive method. Related literature talks mostly about acquiring textual data from social media platforms, very few have examined an amalgamation of numerical and textual data to detect mental illness prevalent in online society. During the process of

literature review, we did not come across any relevant study in literature dealing with causal-approach of the mental illness, also literature lacks in the studies which highlights the correlation between the features and mental health issues. Further the datasets used in the studies were mostly tagged to western and European countries. Very few studies have utilized Indian population as samples. Our empirical study is a concrete step toward addressing gaps identified in the related literature and we propose to automate the process of prediction of mental distress to effectively foster the mental health of online society.

Table 2 Items/features collected through crowd-sourced dataset

1. Gender
 2. Age
 3. Work experience
 4. How many social media accounts do you have?
 5. How frequently do you use social media accounts?
 6. How much time do you spend on Social media every day?
 7. How frequently do you post on Social media?
 8. What do you use social media for?
 - How often during the last year have you... [Spent a lot of time thinking about social media or planned use of social media?](Salience)
 - How often during the last year have you... [felt an urge to use social media more and more?](Tolerance)
 - How often during the last year have you... [Used social media in order to forget about personal problems?](Mood modifications)
 - How often during the last year have you... [Tried to cut down on the use of social media without success?](Relapse)
 - How often during the last year have you... [Become restless or troubled if you have been prohibited from using social media?](Withdrawal)
 - How often during the last year have you... [Used social media so much that it has had a negative impact on your job/studies?](Conflict)
 10. Do you consider yourself addicted to social media?
 11. I'm suffering from Mental health issues
 12. I'm either undergoing "psychiatric medications" or using "antidepressants"
 - [I found it hard to wind down](Stress)
 - [I was aware of dryness of my mouth](Anxiety)
 - [I couldn't seem to experience any positive feeling at all](Depression)
 - [I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion)](Anxiety)
 - [I found it difficult to work up the initiative to do things](Depression)
 - [I tended to over-react to situations](Stress)
 - [I experienced trembling (e.g. in the hands)](Anxiety)
 - [I felt that I was using a lot of nervous energy](Stress)
 - [I was worried about situations in which I might panic and make a fool of myself](Anxiety)
 - [I felt that I had nothing to look forward to](Depression)
 - [I found myself getting agitated](Stress)
 - [I found it difficult to relax](Stress)
 - [I felt down-hearted and blue](Depression)
 - [I was intolerant of anything that kept me from getting on with what I was doing](Stress)
 - [I felt I was close to panic](Anxiety)
 - [I was unable to become enthusiastic about anything](Depression)
 - [I felt I wasn't worth much as a person](Depression)
 - [I felt that I was rather touchy](Stress)
 - [I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat)](Anxiety)
 - [I felt scared without any good reason](Anxiety)
 - [I felt that life was meaningless](Depression)
- Kindly add the last 3 posts you shared in Social media below. Including hashtags if any.

	3.\tWork experience	5.\tHow frequently do you use social media accounts?	6.\tHow much time do you spend on Social media every day?	7.\tHow frequently do you post on Social media?	8.\tWhat do you use social media for?	How often during the last year have you... [Spent a lot of time thinking about social media or planned use of social media?](Salience)	How often during the last year have you... [Felt an urge to use social media more and more?](Tolerance)	How often during the last year have you... [Used social media in order to forget about personal problems?](Mood modifications)	...	240	241	242	243	244	245	246	247	248	249
	2	0	3	0	116	1	2	0	...	0	0	0	0	0	0	0	0	3	9
	3	1	0	1	115	0	0	2	...	146	25	74	52	12	41	116	31	50	40
	4	0	1	2	114	2	1	2	...	0	0	0	0	0	0	0	3	9	147
	3	1	0	1	114	2	0	0	...	15	15	17	96	25	16	25	25	19	33
	3	4	3	1	114	2	2	2	...	0	0	0	0	0	0	0	117	178	148

Fig. 2 Outcome of few encoded items after label encoding

Fig. 5 Embedding Matrix output obtained from our dataset

```
embedding_matrix
array([[ 0.         ,  0.         ,  0.         , ...,  0.         ,
        0.         ,  0.         ],
       [ 0.10343    ,  0.061756   , -0.35646999 , ..., -0.062073    ,
       -0.59389001 , -0.15266     ],
       [ 0.10343    ,  0.061756   , -0.35646999 , ..., -0.062073    ,
       -0.59389001 , -0.15266     ],
       ...,
       [ 0.10343    ,  0.061756   , -0.35646999 , ..., -0.062073    ,
       -0.59389001 , -0.15266     ],
       [ 0.10343    ,  0.061756   , -0.35646999 , ..., -0.062073    ,
       -0.59389001 , -0.15266     ],
       [ 0.10343    ,  0.061756   , -0.35646999 , ..., -0.062073    ,
       -0.59389001 , -0.15266     ]])
```

3.1 Data collection and pre-processing

For the purpose of obtaining a robust automated system, we designed a questionnaire that examine the SNS usage behavior (through Bergen social media addiction scale (Andreassen et al. 2012)), three axes of psychological distress (depression, anxiety, stress) covering physical, social and cognitive experiences (through DASS (Lovibond & Lovibond 1995)) and draw the emotions through logs/posts. Table 2 describes the items of the questionnaire through which we crowd-sourced the data particularly from the Indian population.

The classification performance of a machine learning model is heavily reliant on the good data encoding schemes. Keeping this in view the first step we employed in pre-processing was to turn the categorical data into numerical using label encoder. In Fig. 2, we illustrate the few tokenized data outcome from the set of encoded features. Next the text data comprising of logs/posts is pre-processed by applying Natural Language Processing (NLP) tools using python libraries. The text data are cleaned using a sequence of NLP steps like stripping, removing null, changing entire text into lower case, removing symbols, tokenization, removing stop words, punctuations and lemmatization. In Fig. 3, we present the outcome of the pre-processing steps applied on the logs/posts of our dataset.

3.2 Feature extraction

The efficiency of the text classification lies in the efficient encoding of the unstructured text documents to mathematically computable numerical data while maintaining the semantic and sequence of information in the text. There are different techniques available such as word2vec, word embeddings, glove embeddings etc. to create low dimensional vector space where a large text data is converted to a

2D vector form having an embedding for each word (Kum-nunt and Sornil 2020). In order to mine mental distress cues from the logs, we employed word embeddings and Glove embedding methods of feature extraction. A brief overview of the techniques is discussed below:

3.2.1 Word embeddings

This technique aims to digitize the textual data for the purpose of deep learning by characterizing words or sentences as real number vectors. Basically, semantically similar words are mapped close to each other to retain their meanings. Figure 4, describes the steps to accomplish the word embeddings for the proposed method of detecting mental distress cues from the logs shared by the people in the sample. We have textual input data containing 2500 logs, first step toward transforming these into tensors for the learning purpose of the model is to vectorize the logs corpus by turning each log into sequence of integers (integers are the index of a word in the dictionary). The result is a list of integer sequences with varied sizes (as logs do not have same length). Further, “pad sequence” step is initiated to fill the empty values and truncate the larger sequence to make each vector sequence of uniform length `Max_log_length`. At the end of word embeddings process, the textual data are ready to be joined with the rest of the numerical data in the sample and this word embedding layer become the input to the proposed deep learning model. Further, we also used glove embeddings feature extraction technique to overcome the high dimensionality, sparsity and lack of semantic and syntactic association usually seen in other word embeddings. Glove embeddings is an aggregate global word to word co-occurrence representation showcasing chronological associations to compute relevant features for the classification purpose (Rezaeinia et al. 2019). Figure 5 contains the embedding matrix output obtained from our dataset which

will be fed into the models formulated to predict the occurrence of psychological distress in the sample.

Algorithm 1 General algorithm for all 4 models: prediction of mentally distressed / non- distressed users

BEGIN	
Step-1	Input: Dataset $D=(D_1, D_2, D_3, \dots, D_m)$, $m=2500$, where each $D_i=\{A,S,M,P\}$, $A \in$ Set of general features from the questionnaire (gender, age, no_of_social_media_accounts, frequency, time_spent, no_of_postings etc.) $S \in$ set of Social networking sites addiction features (Saliency, Tolerance, Mood_modification, Withdrawal, Conflict, Relapse) $M \in$ set of three axes of Psychological distress (Depression, Stress, Anxiety) $P \in$ Set of Logs L (Posts on social media platforms)
	Output: Y (Label – 0(No risk)/ 1 (Risk))
Step-2	Encoding of A,S,M using numerical labels
Step-3	Encoding of P - Input for the Word_Embedding Vector : L (A Set of n Logs, $L=L_1, L_2, L_3, \dots, L_n$), $n=2500$
Step-4	For each log L_i in L do $P_i = \text{Word_embeddings}(L_i)$
(Word embeddings for Model 1 and 2/ Glove Embeddings for Model 3 and 4)	
Step-5	Final Vector space $D_i : A_i + S_i + M_i + P_i$ [Each vector comprising of 286 features]
Step-6	Splitting of dataset – Training(samples-1875), Testing(samples-625)
Step-7	
For Model 2 and 4	For each D_i in D do $c_i = \text{CNN}(D_i)$ Feature map C is obtained from MaxPooling layer For each C_i do $Z_i = \text{LSTM}(C_i)$
Step-7	
(For Model 1 and 3)	For each D_i do $Z_i = \text{LSTM}(D_i)$
Step 8	
Dense layer	For each Z_i do $Y_i = \text{sigmoid}(Z_i)$
Step-9	For each Y_i do Output $O_i = \text{compile}(Y_i)$ [loss function: Binary-cross entropy, optimizer:Adam]
Step -10	Print model summary
Step -11	Evaluation and Validation – Test and Predict the mental distressed and non-distressed users.
END	Print confusion matrix matrix, accuracy graph

3.3 Architecture of models for mental distress classification

The classical deep learning models are constructed for prediction of user's vulnerability toward the mental distress due to their online behavior. Algorithm 1 describes the general operating principle of the models. We chose a combination of CNN and LSTM algorithms with word embeddings and global vector embeddings for evaluating the occurrence of psychological distress due to their individual strengths construed from the literature (LeCun and Bengio 1995)

(Hochreiter and Schmidhuber 1997) and pointed out as follows:

- (i) CNN is powerful and has ability to extract as many features as possible from the text.
- (ii) LSTM has a good hold over memorizing sequence of words in a document and it performs fairly better due to its ability to keep relevant words and discard unnecessary words.
- (iii) The goal of word embedding is to reduce dimensionality and use a word to predict the similar words around it thus capturing inter-word semantics. This benefits the computational and generalization power of drawing relevant features.
- (iv) Global vector embedding as the word suggests produces global corpus statistics of the text. The representation of text captures word-word statistics or co-occurrence probabilities of words thus enable efficient text classification.

In the following sub-sections, we give details of the basic mathematical background of CNN and LSTM algorithms and describe step by step learning process of deep learning models determined to resolve the design of detecting mental distress risk.

3.3.1 CNN

Convolutional Neural Network is a type of Artificial Neural Networks which simulates the structure of the human brain neurons for the computational purposes. Generally the neural network architecture is composed of input, output and hidden layers. The hidden layers are connected to each other by weighted links. Mathematical activation functions like sigmoid, hyperbolic Tangent Function (Tanh), Rectified Linear Unit function (ReLU), are applied in each layer to determine the output of each layer. In literature, sigmoid is used for binary classification and Tanh provide better outcomes for multi-layer neural networks (Nwankpa et al. 2018). However, activations in sigmoid ranges between $[0, 1]$ easily saturates and falls almost zero blocking the information flow and tanh performs slightly better with output range between $[-1, 1]$ and is zero-centered. In view of this, ReLU is more popular as its activation threshold is 0 when input becomes less than zero resulting in fast convergence and better performance (L. Zhang et al. 2018).

CNN aims to extracts relevant features through the following layers:

- (i) Convolution Layer.
The central idea in this layer is the sliding or convolving a pre-determined matrix of weights across

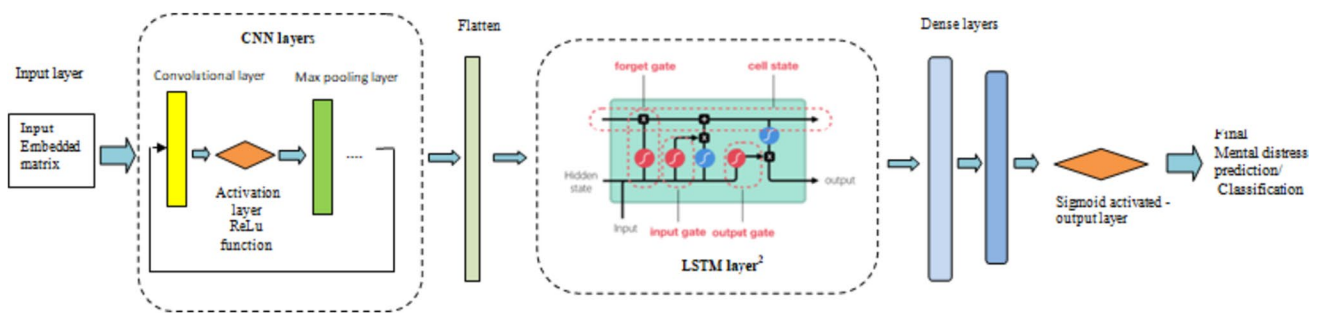


Fig. 6 CNN-LSTM hybrid Architecture for prediction of Mental distress

the input vector space by one step also called stride capturing the relevant features. The weight matrix also called filters is composed of an activation function which decides whether the input to the next hidden layer is relevant or not in the process of prediction. After each filter a layer of max pooling is applied to reduce the vector dimension and the result of all max pooling layers are joined to build input for next neural network in sequence i.e., LSTM in the proposed approach.

(ii) Activation Layer.

As discussed earlier, activation functions are applied in neural networks to transform it to a non-linear network to make it capable to learn and perform complex tasks. However CNNs applies ReLU activation function reducing non-linearity of the network by replacing each negative output by 0, thus unnecessary features are retained which helps to improve prediction accuracy of the models.

(iii) Regularization.

For the purpose of the better generalization and to improve the accuracy of the models on the unseen data, regularization technique is applied. This technique is managed through functions to systematize complex neural network to avoid over-fitting that affects the performance of the models. We use drop-out technique of regularization which penalizes larger weights in order to achieve optimum results.

(iv) Optimization.

During the training phase of deep learning algorithms model parameters like weights, bias and learning rate values are adjusted to reduce the losses and calculated appropriately with optimum values across iterations. There are various optimization approaches like Stochastic Gradient Descent, Adaptive Moment Estimation (Adam), Adagrad etc. We employed Adam optimizer to achieve efficient prediction of mental distress.

Further, we discuss the mathematical concept underlying the CNN feature extraction technique and the description is based on the work presented in (Ghourabi et al. 2020).

Let $Z_i \in D^k$ represents the k-dimensional input vector corresponding to the ith sample. Let $Z \in D^{n \times k}$ denote input vector space, where n is the length of each vector in vector space. For each position j in the vector, let us consider a window vector w_j with p consecutive vector, represented as:

$$w_j = [z_j, z_{j+1}, \dots, z_{j+p-1}] \tag{1}$$

A convolution operation involves a filter $r \in L^{p \times k}$, which is applied to the window w to produce a new feature map $c \in L^{n-p+1}$. Each feature element c_j for window vector w_j is calculated as follows:

$$c_j = f(w_j \odot r + b) \tag{2}$$

where \odot is element-wise product, b is a bias term and f is a nonlinear function, in our case, we used Rectified linear unit (ReLU) as nonlinear function which is generally defined as $f(y) = \max(0, y)$ i.e., it returns y if the value is positive, elsewhere it returns 0.

The convolutional filter is convoluted on each window of input vector to generate a feature map by applying (3).

$$C = [c_1, c_2, c_3, \dots, c_{n-p+1}] \text{ where } c \text{ belongs to } L^{n-p+1} \tag{3}$$

A Max pooling layer is added to reduce the high-level dimensionality of vector generated by the convolution operation. This layer reduces the dimension by selecting only important information and removing weak activation information.

The CNN function mentioned in Algorithm 1 works on the above-mentioned principle to extract optimized important features from the dataset and provide input to the LSTM layer of our proposed mental distress model.

Table 3 Hyper-parameters set up for the deep neural network models

Hyper-parameter	Selected values	Different values tested	Description
Epochs	5–10	2,3,4,5,6,7,8,9,10,11,12	One complete pass of training data in the algorithm
Batch size	64	8,16,32,64,128	Specific number of samples forms batch and model parameters are updated after each batch
Kernel/ filter	128	64,128,256	2D/3D array of weights
Embedding dimension	300	100,200,300	Low dimensional space into which high dimensional vectors can be translated
Learning rate	0.001	0.01,0.001,0.0001,1,0	During tuning of parameters it is the Step size at each iteration to move toward optimization or to minimize loss function
optimizer	Adam	Adagrad, Adam, RMSprop, SGD,	A function that updates the attributes weights and learning rate of neural network
Max_word	250	100,200,250	Maximum number of words in the text
Drop out	0.2	0.1,0.2,0.3,0.4,0.5	Helps in model regularization by dropping neurons in a neural network

Table 4 The system configuration utilized for formation and realization of the proposed models

System specifications	Configuration details
Operating system	Windows,Linux UBUNTU 14.0
Python	3.7
Python library package for stacking	Pandas, Numpy, Sklearn And Vecstack
Python library package for deep learning neural network	Pandas, Numpy, Sklearn, Vecstack Keras and hyperas

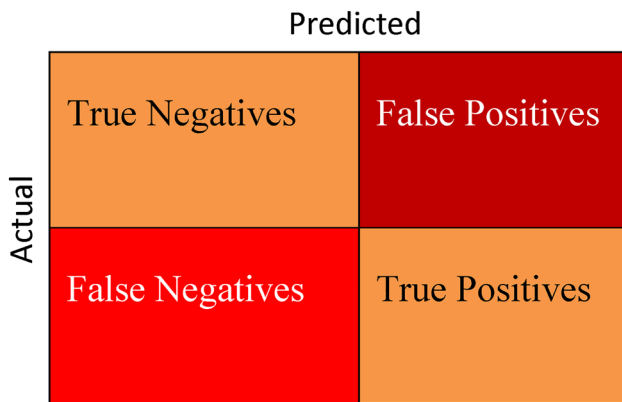


Fig. 7 Confusion matrix

3.3.2 LSTM

Although CNN has useful and efficient qualities of mining out relevant features from the data. However, it is unable to link the current information with the past information thus long-term dependencies in the data is ignored. To overcome this we exploit another deep learning method i.e., LSTM which is an extension of Recurrent Neural Network with a memory to hold over certain pattern that is important to keep during the learning process. The

mathematical architecture of LSTM can be understood by four components consisting of a memory cell c_t and three gates namely input gate i_t , output gate o_t and forget gate f_t . During the learning process these units are repeated for each time step (t) and regulates the flow of information inside the model. To update the memory cell (q_t and c_t) and the hidden state h_t , these gates work collectively and the transition functions used in the process is mentioned as (4) (Ghourabi et al. 2020):

$$\begin{aligned}
 i_t &= \sigma(W_i \cdot [h_{t-1} + b_i]) \\
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 q_t &= \tanh(W_q \cdot [h_{t-1}, x_t] + b_q) \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\
 c_t &= f_t \odot c_{t-1} + i_t \odot q_t \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned}
 \tag{4}$$

where x_t is the input vector for LSTM unit, σ is the sigmoid function, \tanh denotes hyperbolic tangent function, operator \odot signify element wise product, W and b denotes weight matrix and bias parameters respectively that are learned during training of model. We used output vector dimensions as 200 and 100 in two LSTM layers with dropout as 0.2 as regularization parameter to avoid the over-fitting of the model. These functions are executed directly after the Max

Table 5 Description of the key performance metrics—exploited to evaluate the mental distress models

Performance metric	Description	Calculated as	Range of values
Accuracy	Indicates the number of correctly classified instances over the total number of instances in the data	$Accuracy = \frac{TP + TN}{N}$ $(N = TP + TN + FP + FN)$	0 to 1 (0 is worst, 1 is best)
Precision	Indicates ability of a classifier to not classify true negative observation as positive	$precision = \frac{TP}{TP+FP}$	
Recall	Indicates ability of a classifier to find positive observations in the dataset	$recall = \frac{TP}{TP+FN}$	
F1 score	It is an overall measure of a model’s accuracy that combines precision and recall i.e., Harmonic mean of the average precision and recall values of the classification	$F1 = 2 * \frac{precision*recall}{precision+recall}$	

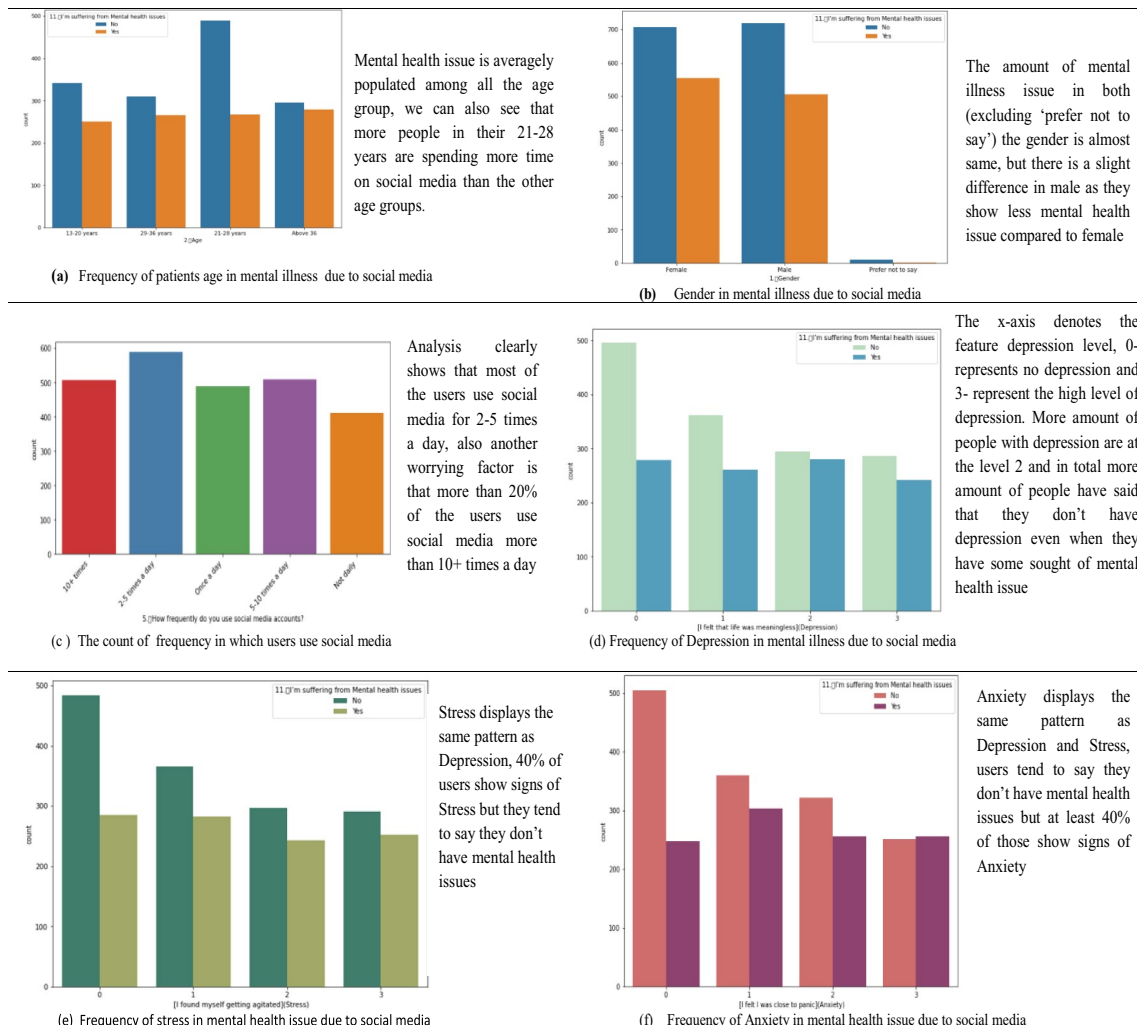


Fig. 8 Illustration of Correlation of variables with mental health risk of users due to their engagement with social networking sites (a) association of age with mental illness (b) Gender relation with mental illness (c) Frequency of social Media usage (d) Frequency of depres-

sion incidence in users with mental illness (e) Frequency of Stress incidence in users with mental illness (f) Frequency of Anxiety incidence in users with mental illness

Table 6 Word embeddings + LSTM model performance outcome along with hyper-parameters set-up

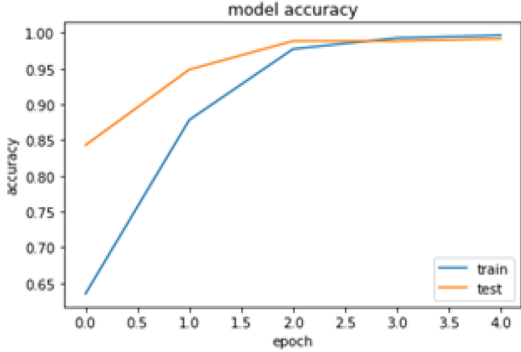
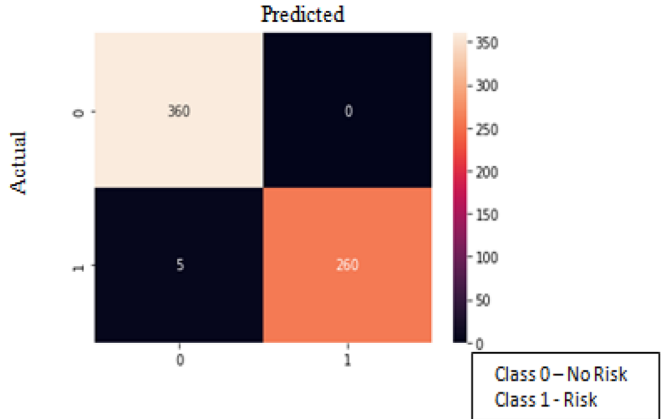
Description and performance outcome of Model 1 (Word Embeddings + LSTM)																
<p>Model Hyper-parameter set up</p> <p>LSTM Dimensionality of output space vector-200,100 Drop out-0.2</p> <p>Dense layer output vector dimension-2 Activation- sigmoid</p> <p>Compile Loss- Binary- cross entropy Optimizer- Adam Epochs-5, Batch size-64, Learning rate-0.001</p>	<p>Model Summary Model: "sequential"</p> <table border="1"> <thead> <tr> <th>Layer (type)</th> <th>Output Shape</th> <th>Param #</th> </tr> </thead> <tbody> <tr> <td>embedding (Embedding)</td> <td>(None, 286, 100)</td> <td>28000</td> </tr> <tr> <td>lstm (LSTM)</td> <td>(None, 286, 200)</td> <td>240800</td> </tr> <tr> <td>lstm_1 (LSTM)</td> <td>(None, 100)</td> <td>120400</td> </tr> <tr> <td>dense (Dense)</td> <td>(None, 2)</td> <td>202</td> </tr> </tbody> </table> <p>Total params: 381,402 Trainable params: 381,402 Non-trainable params: 0</p>	Layer (type)	Output Shape	Param #	embedding (Embedding)	(None, 286, 100)	28000	lstm (LSTM)	(None, 286, 200)	240800	lstm_1 (LSTM)	(None, 100)	120400	dense (Dense)	(None, 2)	202
Layer (type)	Output Shape	Param #														
embedding (Embedding)	(None, 286, 100)	28000														
lstm (LSTM)	(None, 286, 200)	240800														
lstm_1 (LSTM)	(None, 100)	120400														
dense (Dense)	(None, 2)	202														
<p>Training and Testing accuracy graph</p>	<p>Confusion Matrix</p>															

With respect to gender, occurrence of mental health issue is slightly higher in females as compared to males. The analysis of count of frequency in which users use social media shows 20% of users use social media more than 10+ times in a day. Also viewing the analysis of psychological distress axes we observed that 40% of users have experienced the episodes of mental distress (depression, stress, anxiety) but they tend to ignore it and say that ‘they don’t have any mental health issues’. Thus, in order to help such individuals who may go through serious detrimental health issues later on, we can say that our study of automatic detection of mental distress is important and the need of the hour to uphold the wellness of the online community.

Further in the next step, we examined the logs/posts of the dataset for the purpose of finding the articulation of the comments i.e., what posts really speak about a person’s wellbeing. The term ‘affect’ in psychology describes the expressions of mood and emotions which plays crucial role in indentifying the well-being of a person (Chen et al. 2020a, b) (Silvera et al. 2008). The affective expressions in virtual community has come into sight for making inferences about person’s mental health status (De Choudhury et al. 2013)

(Bazarova et al. 2015). In view of above we formulate the association between frequency of affective words in logs and mental well being by illustrating the word cloud of mental distressed and non-distressed posts in Fig. 9. The size of the words depicts the frequency of the words in the logs and it is visually evident in Fig. 9a, affective words denoting mental illness are frequently present in the logs of distressed group in contrast to the normal words in the logs of the non-distressed group in Fig. 9b. Next we discuss the performance of four deep neural network models constructed using word embeddings and glove embeddings. For the validation purpose we trained our models with 70% training dataset and 30% testing dataset. The first model we realized consisted of three layers—word embedding layer and two LSTM layers. By adding one more dropout layer to this combination we got 98.72% accuracy. Second model consisted of five layers—word embedding layer, a CNN layer, a LSTM layer, Max-pooling and dense layer, achieved 99.20% accuracy. Third model employed pre-trained glove vectors with LSTM capabilities to predict the mental distress in the population with 98.72% accuracy. Finally in the fourth model glove embeddings is employed as the first layer after that

Table 7 Word embeddings + CNN-LSTM model performance outcome along with hyper-parameters set-up

Description and performance outcome of Model 2 (Word Embeddings + CNN- LSTM)																			
Model Hyper-parameter set up CNN Filters=32, Kernel size=3, activation='relu', padding='same', pool size=2 LSTM output vector dimension - 100 Drop out - 0.2 Dense layer output vector dimension-2 Activation - sigmoid Compile Loss- Binary -cross entropy Optimizer-Adam Epochs-5, Batch size-64, Learningrate-0.001	Model Summary Model: "sequential_1" <table border="1"> <thead> <tr> <th>Layer (type)</th> <th>Output Shape</th> <th>Param #</th> </tr> </thead> <tbody> <tr> <td>embedding_1 (Embedding)</td> <td>(None, 286, 100)</td> <td>28000</td> </tr> <tr> <td>conv1d (Conv1D)</td> <td>(None, 286, 32)</td> <td>9632</td> </tr> <tr> <td>max_pooling1d (MaxPooling1D)</td> <td>(None, 143, 32)</td> <td>0</td> </tr> <tr> <td>lstm_2 (LSTM)</td> <td>(None, 100)</td> <td>53200</td> </tr> <tr> <td>dense_1 (Dense)</td> <td>(None, 2)</td> <td>202</td> </tr> </tbody> </table> Total params: 83,034 Trainable params: 83,034 Non-trainable params: 0	Layer (type)	Output Shape	Param #	embedding_1 (Embedding)	(None, 286, 100)	28000	conv1d (Conv1D)	(None, 286, 32)	9632	max_pooling1d (MaxPooling1D)	(None, 143, 32)	0	lstm_2 (LSTM)	(None, 100)	53200	dense_1 (Dense)	(None, 2)	202
Layer (type)	Output Shape	Param #																	
embedding_1 (Embedding)	(None, 286, 100)	28000																	
conv1d (Conv1D)	(None, 286, 32)	9632																	
max_pooling1d (MaxPooling1D)	(None, 143, 32)	0																	
lstm_2 (LSTM)	(None, 100)	53200																	
dense_1 (Dense)	(None, 2)	202																	
Training and Testing accuracy graph 	Confusion Matrix 																		

Convolutional neural network is added followed by max-pooling to extract only important information and then pass the outcome to the LSTM layer. The testing accuracy we got for this model is around 98.40%. The performance result of all four classification models applied on the dataset along with the hyper parameter setup is presented in Tables 6, 7, 8 and 9 respectively.

Furthermore, we present the comparison of all four models with respect to performance metrics like precision, recall and F1 score in Table 10. We accomplished 100% precision for word embeddings combined with CNN-LSTM capabilities. Precision performance metric plays an important role in finding the utility of the models in health sector problems as it denotes the model has produced zero false positives thus evades the trouble and needless intervention which may be caused due to the false alarm. The higher F1 score from the models can be explained by the use of the ReLu activation function having the ability to obstruct negative activation thus returns

improved performance as the number of parameters to be learned is reduced.

Finally, in Table 11, we present a comparison of our study with the other related recent studies in literature and found that our hybrid deep neural network model outperformed others in accuracy. We also infer that previous studies are based on analysis of only textual data in contrast to our study where we focused on textual data and causal factors (Social networking site addiction, Depression, Anxiety and Stress state) as well. This explains the novelty of our approach and the empirical outcomes in this study using a novel feature construct is a better step toward the detection and assessment of mental distress risk helping people to cope-up with their problems.

Additionally, first we used crowd-sourced textual data for tuning BERT, a state-of-the-art advanced NLP model. BERT makes contextualized word vectors which makes inference easier based on the context learning (Chandra and Kulkarni 2022). Our dataset result shows an enhanced

Table 8 Glove embeddings + LSTM model performance outcome along with hyper-parameters set-up

Description and performance outcome of Model 3 (Glove Embeddings + LSTM)																
<p>Model Hyper-parameter set up</p> <p>LSTM Dimensionality of output space vector-200,100 Drop out-0.2</p> <p>Dense layer output vector dimension-2 Activation- sigmoid</p> <p>Compile Loss- Binary- cross entropy Optimizer- Adam Epochs-5, Batch size-64, Learning rate-0.001</p>	<p>Model Summary Model: "sequential_2"</p> <table border="1"> <thead> <tr> <th>Layer (type)</th> <th>Output Shape</th> <th>Param #</th> </tr> </thead> <tbody> <tr> <td>embedding_2 (Embedding)</td> <td>(None, 286, 300)</td> <td>2627100</td> </tr> <tr> <td>lstm_3 (LSTM)</td> <td>(None, 286, 200)</td> <td>400800</td> </tr> <tr> <td>lstm_4 (LSTM)</td> <td>(None, 100)</td> <td>120400</td> </tr> <tr> <td>dense_2 (Dense)</td> <td>(None, 2)</td> <td>202</td> </tr> </tbody> </table> <p>Total params: 3,148,502 Trainable params: 3,148,502 Non-trainable params: 0</p>	Layer (type)	Output Shape	Param #	embedding_2 (Embedding)	(None, 286, 300)	2627100	lstm_3 (LSTM)	(None, 286, 200)	400800	lstm_4 (LSTM)	(None, 100)	120400	dense_2 (Dense)	(None, 2)	202
Layer (type)	Output Shape	Param #														
embedding_2 (Embedding)	(None, 286, 300)	2627100														
lstm_3 (LSTM)	(None, 286, 200)	400800														
lstm_4 (LSTM)	(None, 100)	120400														
dense_2 (Dense)	(None, 2)	202														
<p>Training and Testing accuracy graph</p>	<p>Confusion Matrix</p> <table border="1"> <thead> <tr> <th colspan="2" rowspan="2"></th> <th colspan="2">Predicted</th> </tr> <tr> <th>0</th> <th>1</th> </tr> </thead> <tbody> <tr> <th rowspan="2">Actual</th> <th>0</th> <td>359</td> <td>1</td> </tr> <tr> <th>1</th> <td>7</td> <td>258</td> </tr> </tbody> </table> <p>Class 0 - No Risk Class 1 - Risk</p>			Predicted		0	1	Actual	0	359	1	1	7	258		
				Predicted												
		0	1													
Actual	0	359	1													
	1	7	258													

performance as we achieved 100% accuracy, precision, recall and F1 score. Next, to validate our results, we used 3000 tweets mined from Indian twitter for fine tuning of BERT model and achieved 96% accuracy. The outcome of both these experiments is highlighted in Table 11. The results fortify our objective of exploring the logs for mental-distress cues.

5 Conclusion

Our research contributed toward the identification of mental distress prevalent in online community by experimenting with a hybrid AI model built on novel feature construct consisting of numerical simulation combined with textual analysis which is not present in any related type of study found in literature currently. The advantages of two efficient deep learning algorithms namely convolutional neural network algorithm and Long Short-Term Memory algorithm are hybridized to deliver an efficient automated mental

health detection scale. The dataset used to train and test the performance of the models was crowd-sourced and can be considered as benchmark for further research in this area. In order to extract mental cues from text, Glove embeddings layer was pooled with a CNN+LSTM hybrid neural network model and after a fine tuning of hyper-parameters, we achieved a 99.20% accuracy and 100% precision outperforming any models found in literature currently. With zero false positives we reached a greater degree of clarity and efficiency with respect to the disturbance and unnecessary treatment a person may face due to false alarm. Additionally, for validation purpose we exploited a different ground truth dataset from general Indian twitter and employed a modern NLP model called BERT.

We illustrated the association of mental health issues with features like age, gender and frequency of social media usage. Also presented graphically the relationship of depression, stress and anxiety quotient with respect to the mental health state of the people participated in the study. We found that people experience depression, stress and anxiety but prefer not to disclose. It becomes highly

Table 9 Glove embeddings + CNN-LSTM model performance outcome along with hyper-parameters set-up

Description and performance outcome of Model 4 (Glove Embeddings + CNN-LSTM)																									
Model Hyper-parameter set up CNN Filters=128,Kernel size=3, activation='relu', padding='same',pool size=2 LSTM output vector dimension-100 Drop out-0.2 Dense layer output vector dimension-2 Activation- sigmoid Compile Loss- Binary- cross entropy Optimizer- Adam Epochs-5, Batch size-64, Learningrate-0.001	Model Summary Model: "sequential_3" <table border="1"> <thead> <tr> <th>Layer (type)</th> <th>Output Shape</th> <th>Param #</th> </tr> </thead> <tbody> <tr> <td>embedding_3 (Embedding)</td> <td>(None, 286, 300)</td> <td>2627100</td> </tr> <tr> <td>conv1d_1 (Conv1D)</td> <td>(None, 286, 128)</td> <td>115328</td> </tr> <tr> <td>max_pooling1d_1 (MaxPooling1D)</td> <td>(None, 143, 128)</td> <td>0</td> </tr> <tr> <td>lstm_5 (LSTM)</td> <td>(None, 100)</td> <td>91600</td> </tr> <tr> <td>dense_3 (Dense)</td> <td>(None, 2)</td> <td>202</td> </tr> </tbody> </table> Total params: 2,834,230 Trainable params: 2,834,230 Non-trainable params: 0	Layer (type)	Output Shape	Param #	embedding_3 (Embedding)	(None, 286, 300)	2627100	conv1d_1 (Conv1D)	(None, 286, 128)	115328	max_pooling1d_1 (MaxPooling1D)	(None, 143, 128)	0	lstm_5 (LSTM)	(None, 100)	91600	dense_3 (Dense)	(None, 2)	202						
Layer (type)	Output Shape	Param #																							
embedding_3 (Embedding)	(None, 286, 300)	2627100																							
conv1d_1 (Conv1D)	(None, 286, 128)	115328																							
max_pooling1d_1 (MaxPooling1D)	(None, 143, 128)	0																							
lstm_5 (LSTM)	(None, 100)	91600																							
dense_3 (Dense)	(None, 2)	202																							
Training and Testing accuracy graph 	Confusion Matrix <table border="1"> <thead> <tr> <th colspan="2"></th> <th colspan="2">Predicted</th> <th></th> </tr> <tr> <th colspan="2"></th> <th>0</th> <th>1</th> <th></th> </tr> </thead> <tbody> <tr> <th rowspan="2">Actual</th> <th>0</th> <td>357</td> <td>3</td> <td></td> </tr> <tr> <th>1</th> <td>7</td> <td>258</td> <td></td> </tr> <tr> <td></td> <td></td> <td>0</td> <td>1</td> <td></td> </tr> </tbody> </table> <p>Class 0 – No Risk Class 1 - Risk</p>			Predicted					0	1		Actual	0	357	3		1	7	258				0	1	
		Predicted																							
		0	1																						
Actual	0	357	3																						
	1	7	258																						
		0	1																						

Table 10 Mental distress detection results (*Best performing model)

		Performance metrics			
		Accuracy (%)	Precision	Recall	F1-Score
Models for predicting mental distress	Embedding layer +LSTM	98.72	0.99	0.97	0.979
	Embedding Layer+ CNN-LSTM*	99.20	1	0.98	0.989
	GloVe Embeddings + LSTM	98.72	0.99	0.97	0.979
	GloVe Embeddings + CNN-LSTM	98.40	0.98	0.97	0.974

important to develop robust automated system which can help such type of people in predicting their mental state thus help them to move on to timely healthy interventions. Our hybrid model is a tangible step toward it and sets up an inexpensive method of collection of experiences of online community to address the challenges of timely detection of mental distress.

The learning from the study brings out a feasible tool wherein apt natural language processing methods are coupled with deep learning capabilities to predict the SNS users' likeliness of psychological distress risk. However, the study has a limitation in generalizing the outcome attained

to the virtual community as a whole, due to the common method bias factor which is unavoidable when data is collected through a self-report questionnaire. Another limitation this study has is the shorter time period of collection of samples. The present cross-sectional study investigated mental health issue by collecting data from a population in a particular time frame, in future we intend to design longitudinal methodology which may be beneficial in establishing a better trait of mental health issues.

As a future scope, we propose to widen the feature construct by capturing the facial experiences and mapping the changes in the brain through an inexpensive method. There

Table 11 Comparison of the proposed study with other related recent studies in literature

Paper title	Authors/Year	Dataset	Type of mental disorder	DL/ML techniques	Result	Feature extraction techniques	Sample size
Characterization of mental health conditions in social media using Informed Deep Learning	Gkothis et al. (2017)	Reddit posts	11 mental themes like BPD, bipolar, schizophrenia, anxiety, depression etc	CNN,FF,Linear, SVM	Accuracy: CNN-91.08% FF-90.78% Linear-86.01% SVM-85.87%	Word Embeddings	538,245 Posts related to 11 mental themes, 476,388 non-mental health post
Detection of Depression in Thai Social Media Messages using Deep Learning	Kumnunt and Sornil (2020)	Thai social media messages	Depression	CNN-LSTM	Accuracy-83.1%	Word2Vec	5,283 depression posts 26,415 control posts
An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM	Kour et al. (2022)	Twitter API	Depression	CNN + biLSTM	Accuracy: 94.28%	Word Embeddings	58,810 samples
Detection of Types of Mental Illness through the Social Network Using Ensembled Deep Learning Model	Nasrullah et al. (2022)	Extracted from Reddit	Anxiety, bipolar, dementia, psychotic	CNN + biLSTM	Accuracy-92%	Continuous Bag of words	301,746 records
Deep Learning for Depression Detection from Textual Data	Amanat A et al. (2022)	Tweets from Kaggle	Depression	LSTM-RNN	Accuracy-99%	One-hot,PCA	4000 tweets
A Novel Text Mining Approach for Mental Health Prediction Using Bi-LSTM and BERT Model	Zeberga et al. (2022)	Twitter and Reddit	Depression, Anxiety	Bi-LSTM and BERT	Accuracy- 98%	Word2Vec, BERT	Reddit -1,55,000 posts Twitter- 40,000
Proposed study		Crowd-sourceddata	Psychological distress	CNN-LSTM	Accuracy-99,20%	Embeddings, GloVe embeddings	2500 logs
		Crowd-sourced Logs		BERT	Accuracy-100%	BERT Vectors	2500 logs
		Indian Twitter tweets		BERT	Accuracy-96%	BERT Vectors	3000 Tweets

are studies in literature related to facial and brain mapping feature analysis but collection of such type of data is expensive and not possible practically in day-to-day life. Finally, we suggest as our future perspective to keep enhancing the automatic detection process of mental distress in order to uphold the online community health.

Authors contributions The manuscript was designed and written together by all the three authors Dr. MKA, Dr. JS and AS. The experiment in the study was performed by the author Anju Singh, analysis and results were reviewed by the authors Dr. MKA and Dr. JS. All the authors contributed equally in this study and vouch for the results reported and protocol adherence during the study. All the three authors edited and approved the manuscript.

Funding No funds, grants, or other support were received for conducting this study and preparation of the manuscript.

Data availability Data will be made available upon reasonable request for further study in this area of research within the limitations of the informed consent by the corresponding author upon acceptance from the co-author.

Declarations

Conflict of interest There is no competing interest's disclosure relevant to this study and related to any of the authors of this manuscript.

References

- Abd Rahman R, Omar K, Mohd Noah SA, Mohd Danuri MSN (2018) A survey on mental health detection in online social network. *Int J Adv Sci Eng Inform Technol* 8:1431. <https://doi.org/10.18517/ijaseit.8.4-2.6830>
- Ahmed S, Abu Yousuf M, Monowar MM, Hamid A, Allassafi MO (2023) Taking all the factors we need: a multimodal depression classification with uncertainty approximation. *IEEE Access* 11:99847–99861
- Ahmed A, Yousuf M (2021) Sentiment analysis on bangla text using long short-term memory (LSTM) recurrent neural network. pp 181–192. doi https://doi.org/10.1007/978-981-4673-4_16
- Amanat A, Rizwan M, Javed AR, Abdelhaq M, Alsaqour R, Pandya S, Uddin M (2022) Deep learning for depression detection from textual data. *Electronics* 11(5):676. <https://doi.org/10.3390/electronics11050676>
- Andreassen CS, Torsheim T, Brunborg GS, Pallesen S (2012) Development of a facebook addiction scale. *Psychol Rep* 110(2):501–517. <https://doi.org/10.2466/02.09.18.PR0.110.2.501-517>
- Bazarova NN, Choi YH, Schwanda Sosik V, Cosley D, Whitlock J (2015) Social sharing of emotions on facebook: channel differences, satisfaction, and replies. In: proceedings of the 18th ACM conference on computer supported cooperative work & social computing. pp 154–164. <https://doi.org/10.1145/2675133.2675297>
- Belfin RV, Kanaga EGM, Kundu S (2020) Application of machine learning in the social network. Recent advances in hybrid metaheuristics for data clustering. Wiley, New Jersey, pp 61–83. <https://doi.org/10.1002/9781119551621.ch4>
- Chancellor S, De Choudhury M (2020) Methods in predictive techniques for mental health status on social media: a critical review. *Npj Dig Med* 3(1):43. <https://doi.org/10.1038/s41746-020-0233-7>
- Chandra R, Kulkarni V (2022) Semantic and sentiment analysis of selected bhagavad gita translations using BERT-based language framework. *IEEE Access* 10:21291–21315. <https://doi.org/10.1109/ACCESS.2022.3152266>
- Chen LL, Magdy W, Wolters MK (2020) The effect of user psychology on the content of social media posts: originality and transitions matter. *Front Psychol* 11:1024. <https://doi.org/10.3389/fpsyg.2020.00526>
- Chen L, Xia C, Sun H (2020b) Recent advances of deep learning in psychiatric disorders. *Precis Clin Med* 3(3):202–213. <https://doi.org/10.1093/pcmedi/pbaa029>
- Cho G, Yim J, Choi Y, Ko J, Lee S-H (2019) Review of machine learning algorithms for diagnosing mental illness. *Psychi Invest* 16(4):262–269. <https://doi.org/10.30773/pi.2018.12.21.2>
- Cong Q, Feng Z, Li F, Xiang Y, Rao G, Tao C (2018) X-A-BiLSTM: a deep learning approach for depression detection in imbalanced data. In: 2018 IEEE international conference on bioinformatics and biomedicine (BIBM). <https://doi.org/10.1109/BIBM.2018.8621230>
- Coppersmith G, Leary R, Crutchley P, Fine A (2018) Natural language processing of social media as screening for suicide risk. *Biomedical Informatics Insights* 10:1178222618792860. <https://doi.org/10.1177/1178222618792860>
- D'Alfonso S (2020) AI in mental health. *Curr Opin Psychol* 36:112–117. <https://doi.org/10.1016/j.copsyc.2020.04.005>
- De Choudhury M, Counts S, Horvitz E (2013) Social media as a measurement tool of depression in populations. *Proc Ann ACM Web Sci Conf WebSci*. 13:47–56. <https://doi.org/10.1145/2464464.2464480>
- Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J (2018) Mental health monitoring with multimodal sensing and machine learning: a survey. *Pervasive Mob Comput* 51:1–26. <https://doi.org/10.1016/j.pmcj.2018.09.003>
- Ghosh T, Palash MIA, Yousuf MA, Hamid MA, Monowar MM, Allassafi MO (2023) A robust distributed deep learning approach to detect alzheimer's disease from MRI images. *Mathematics* 11(12):12. <https://doi.org/10.3390/math11122633>
- Ghourabi A, Mahmood MA, Alzubi QM (2020) A hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages. *Future Internet* 12(9):156. <https://doi.org/10.3390/fi12090156>
- Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, Dutta R (2017) Characterization of mental health conditions in social media using informed deep learning. *Sci Rep* 7(1):45141. <https://doi.org/10.1038/srep45141>
- Glaz AL, Haralambous Y, Kim-Dufor D-H, Lenca P, Billot R, Ryan TC, Marsh J, DeVlyder J, Walter M, Berrouguet S, Lemey C (2021) Machine learning and natural language processing in mental health: systematic review. *J Med Internet Res* 23(5):e15708. <https://doi.org/10.2196/15708>
- Haque UM, Kabir E, Khanam R (2021) Detection of child depression using machine learning methods. *PLoS ONE* 16(12):e0261131. <https://doi.org/10.1371/journal.pone.0261131>
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Husseini Orabi A, Buddhitha P, Husseini Orabi M, Inkpen D (2018) Deep learning for depression detection of twitter users. In: proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic. pp 88–97. doi <https://doi.org/10.18653/v1/W18-0609>
- Iannotti RJ, Janssen I, Haug E, Kololo H, Annaheim B, Borraccino A (2009) Interrelationships of adolescent physical activity, screen-based sedentary behavior, and social and psychological health. *Int J Public Health* 54(Suppl 2):191–198. <https://doi.org/10.1007/s00038-009-5410-z>

- Jo H, Kim S-M, Ryu J (2017) What we really want to find by Sentiment Analysis: the relationship between computational models and psychological state. *ArXiv*.
- Kanaan R, Haidar B, Kilany R (2021) Detecting mental disorders through social media content. In: 2021 IEEE 3rd international multidisciplinary conference on engineering technology (IMCET), 23–28. <https://doi.org/10.1109/IMCET53404.2021.9665555>
- Karim F, Oyewande AA, Abdalla LF, Ehsanullah RC, Khan S (2020) Social media use and its connection to mental health: a systematic review. *Cureus* 12(6):15. <https://doi.org/10.7759/cureus.8627>
- Kim J, Lee J, Park E, Han J (2020) A deep learning model for detecting mental illness from user content on social media. *Sci Rep* 10:11846. <https://doi.org/10.1038/s41598-020-68764-y>
- Kim J, Lee D, Park E (2021) Machine learning for mental health in social media: bibliometric study. *J Med Internet Res* 23(3):e24870. <https://doi.org/10.2196/24870>
- Kour H, Gupta MK (2022) An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. *Multim Tools Appl*. <https://doi.org/10.1007/s11042-022-12648-y>
- Kumnunt B, Sornil O (2020) Detection of depression in thai social media messages using deep learning. In: proceedings of the 1st international conference on deep learning theory and applications, pp 111–118. doi <https://doi.org/10.5220/0009970501110118>
- LeCun Y, Bengio Y (1995) Convolutional networks for images, speech, and time series. In: Arbib MA (ed) *Handbook of brain theory and neural networks*. MIT Press, Cambridge, p 3361
- Lin H, Jia J, Qiu J, Zhang Y, Shen G, Xie L, Tang J, Feng L, Chua T-S (2017) Detecting stress based on social interactions in social networks. *IEEE Trans Knowl Data Eng* 29(9):1820–1833. <https://doi.org/10.1109/TKDE.2017.2686382>
- Lovibond PF, Lovibond SH (1995) The structure of negative emotional states: comparison of the depression anxiety stress scales (dass) with the beck depression and anxiety inventories. *Behav Res Ther* 33(3):335–343. [https://doi.org/10.1016/0005-7967\(94\)00075-u](https://doi.org/10.1016/0005-7967(94)00075-u)
- Mental health in Asia: The numbers. (n.d.). Retrieved October 27, 2021, from <https://www.ourbetterworld.org/series/mental-health/support-toolkit/mental-health-asia-numbers>
- Nwankpa C, Ijomah W, Gachagan A, Marshall S (2018) Activation functions: comparison of trends in practice and research for deep learning (arXiv:1811.03378). arXiv. <http://arxiv.org/abs/1811.03378>
- Rahman RA, Omar K, Mohd Noah SA, Danuri MSNM, Al-Garadi MA (2020) Application of machine learning methods in mental health detection: a systematic review. *IEEE Access* 8:183952–183964. <https://doi.org/10.1109/ACCESS.2020.3029154>
- Ramírez-Cifuentes D, Largeron C, Tissier J, Baeza-Yates R, Freire A (2021) Enhanced word embedding variations for the detection of substance abuse and mental health issues on social media writings. *IEEE Access* 9:130449–130471. <https://doi.org/10.1109/ACCESS.2021.3112102>
- Rezaeinia SM, Rahmani R, Ghodsi A, Veisi H (2019) Sentiment analysis based on improved pre-trained word embeddings. *Expert Syst Appl* 117:139–147. <https://doi.org/10.1016/j.eswa.2018.08.044>
- Sagar R, Pattanayak RD, Chandrasekaran R, Chaudhury PK, Deswal BS, Lenin Singh RK, Malhotra S, Nizamie SH, Panchal BN, Sudhakar TP, Trivedi JK, Varghese M, Prasad J, Chatterji S (2017) Twelve-month prevalence and treatment gap for common mental disorders: findings from a large-scale epidemiological survey in India. *Indian J Psych* 59(1):46–55. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_333_16
- Saqib K, Khan AF, Butt ZA (2021) Machine learning methods for predicting postpartum depression: scoping review. *JMIR Mental Health* 8(11):e29838. <https://doi.org/10.2196/29838>
- Shatte ABR, Hutchinson DM, Teague SJ (2019) Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 49(9):1426–1448. <https://doi.org/10.1017/S0033291719000151>
- Silvera DH, Lavack AM, Kropp F (2008) Impulse buying: the role of affect, social influence, and subjective wellbeing. *J Consum Mark* 25(1):23–33. <https://doi.org/10.1108/07363760810845381>
- Su C (2020) Deep learning in mental health outcome research: a scoping review. *Transl Psych* 26(1):116
- Syaputra RA, Ali R (2022) Chapter Ten—Improving mental health surveillance over Twitter text classification using word embedding techniques. In: Jain S, Pandey K, Jain P, Seng KP (eds) *Artificial intelligence, machine learning, and mental health in pandemics*. Academic Press, NJ, pp 235–258. <https://doi.org/10.1016/B978-0-323-91196-2.00014-4>
- Thieme A, Belgrave D, Doherty G (2020) Machine learning in mental health: a systematic review of the HCI literature to support effective ML system design. *ACM Trans Comput Human Int (TOCHI)*, 27. <https://www.microsoft.com/en-us/research/publication/machine-learning-in-mental-health-a-systematic-review-of-the-hci-literature-to-support-effective-ml-system-design/>
- Zeberga K, Attique M, Shah B, Ali F, Jembre YZ, Chung T-S (2022) A novel text mining approach for mental health prediction using Bi-LSTM and BERT model. *Comput Intell Neurosci* 2022:1–18. <https://doi.org/10.1155/2022/7893775>
- Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: A survey. *WIREs Data Min Knowl Discov* 8(4):e1253. <https://doi.org/10.1002/widm.1253>
- Zhang T, Schoene AM, Ji S, Ananiadou S (2022) Natural language processing applied to mental illness detection: a narrative review. *Npj Dig Med* 5(1):46. <https://doi.org/10.1038/s41746-022-00589-7>
- Zhang Z, Lin W, Liu M, Mahmoud Rady M (2020) Multimodal deep learning framework for mental disorder recognition. pp 344–350. Doi <https://doi.org/10.1109/FG47880.2020.00033>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.