

Sentiment-aware bidirectional long short-term memory model for depression detection from social media texts

Cephas Paul Edward, Karthikeyan Harimoorthy

Department of Computer Science and Engineering, Christ University, Bengaluru, India

Article Info

Article history:

Received Jul 12, 2025

Revised Jul 17, 2026

Accepted Aug 12, 2026

Keywords:

Affective computing

Bidirectional long short-term memory

Deep learning

Depression detection

Sentiment analysis

ABSTRACT

The number of people living with depression is growing at an alarming rate. Depression impairs normal life, ranging from loss of performance in studies or work to being the cause for other health issues either due to lack of care or self-harm and in rare cases, even suicide. Early-detection is crucial for a better course of treatment. Younger generations usually tend to avoid seeking medical help due to the fear of being “judged” but on the other hand share their feelings on social media. There have been various studies in the past to detect depression from social media posts but detection accuracy is limited due to a number of factors such as informal language. Modeling emotional states from text continues to be a core challenge in human-centered artificial intelligence (AI). The sentiment-aware composite attention bidirectional long short-term memory (SCAB-Net) model aids in screening regular social media posts from the ones indicative of depression. The results show that the proposed model outperforms nine baseline models - support vector machine (SVM), long short-term memory (LSTM), bidirectional long short-term memory (BiLSTM), convolutional neural network (CNN), CNN LSTM, CNN BiLSTM, CNN with attention, BiLSTM with attention, CNN bidirectional gated recurrent unit (BiGRU) achieving an accuracy of 89.85%.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Karthikeyan Harimoorthy

Department of Computer Science and Engineering, Christ University

Bengaluru, India

Email: karthikeyan.h@christuniversity.in

1. INTRODUCTION

Depression is a prevalent issue in the current day society. A study by McKinsey shows that younger generations have poor emotional well-being and exhibit behavioral-health challenges. Behavioral health issues are caused by underlying mental health problems such as trauma and stress [1]. A more recent study from Gallup also reinforces the same understanding that mental health issues are prevalent among them. The impact of these mental health issues, such as depression, not only affects the personal well-being and the behavioral well-being but the future workforce is at stake as they are exhibiting higher levels of depression and anxiety [2]–[5]. A survey of 36,552 respondents showed that there is an increase in depressive symptoms in adolescence, which continues into young adulthood [6]. A study by the Rollins School of Public Health conducted in May 2024 states social media has on the other hand, has promoted openness to bring up and discuss mental health problems [7].

Young people are reluctant to seek medical help for depression due to pervasive stigma that manifests in feelings of being denied, isolated, or blamed for their condition. They may perceive mental health struggles

are being dismissed as attention-seeking or as a sign of weakness, leading to self-doubt and internalization of negative beliefs [8], [9]. This generation is particularly sensitive to societal perceptions and may worry about being labeled as “weak” or “crazy” [10]. The younger generations are heavy users of social media and it has a heavy impact on their emotional well-being. They actively share their feelings on social media platforms through posts, photos and videos [11]–[13].

There has been research in the past establishing the ability to leverage social media texts in order to detect event sentiments and mental health disorders such as anxiety, depression leveraging artificial intelligence (AI) and machine learning approaches [14], [15]. Depression causes the affected individuals to manifest subtle and non-linear patterns in speech, textual content and behavior. Detecting such complex and high-dimensional patterns are something which machine learning techniques excel at. For example, posting at odd times of the day, posts that indicate loneliness or sentimentally negative or suicidal content can be detected by leveraging appropriate machine learning or deep learning techniques [16]. This paper takes into consideration the existing work on depression detection via deep learning and affective computing techniques and proposes an improvised model that can learn, adapt and reason over emotional text and hence can be incorporated in intelligent systems for health monitoring. The proposed improvisation is an innovative ensemble method that combines a neural network along with a rule-based classifier that is sensitive to text with depressive indicators that can be further extended to sentiment analysis and affective computing problems.

The paper is structured in the following manner. Section 1 provides a brief introduction to the pressing need and the concepts around detection of depression from social media. Section 2 outlines the related work in this domain in detail. Section 3 describes the architecture, algorithms and details of the novel sentiment-aware composite attention bidirectional long short-term memory (SCAB-Net). Section 4 illustrates the findings, followed by section 5, which covers the conclusion and scope for future work.

2. RELATED WORK

Studies establish that hybrid approaches combining multiple algorithms enhance the accuracy and efficiency of depression detection compared to traditional methods. Long short-term memory (LSTM) and bidirectional long short-term memory (BiLSTM) are the most common neural networks used in this domain. Pretrained language models (PLMs) such as bidirectional encoder representations from transformers (BERT) are commonly used for embeddings. However, there are only a few studies that integrate sentiment cues explicitly [17]. The models are also heavily tied to the language of the posts/tests [18].

An attention-based convolutional neural network (CNN)-BiLSTM combined with feature extraction based on a 1D convolutional layer on embedded vector size 100 was used for detecting depression from social media posts. A bias term was followed by a max pooling layer of size 3 for producing feature maps. Training was performed using a BiLSTM model with 128 hidden states, which helps in summarizing the contextual information of the entire sentence. This was introduced earlier for positive/negative sentiment analysis from reviews and applied for the depression detection domain. This approach achieved an accuracy of 96.7% but the results were limited to posts explicitly mentioning depression and lacks generalizability to texts without explicit mention to depression [19]. Dalal *et al.* [20] performed a qualitative systematic literature review of advancements in depression detection using social media data. The review attempts to list the feature categories from existing work sentiment, emotion, content, psycholinguistic, posting behavior, and user profile. Most of the datasets only consist of textual information. Few datasets can be considered as multi-modal, having hashtags and other information of the user, such as geographic location, age. PLMs such as BERT and domain-specific models such as MentalBERT and PsychBERT exist for feature embeddings and are generally used. While there are numerous features, domain-specific and sentiment features have not been explored widely.

A hierarchical ensemble learning approach leveraged an open access dataset from Kaggle with 7,489 samples (with 1,227 samples suggestive of depression). Text is pre-processed, followed by an ensemble of embedding BERT (sentence level) and global vectors for word representation (GloVe) based textual representations. The dual vectors are then passed into ensemble of classifiers leveraging dense neural network (DNN), text-specific convolutional neural network (TextCNN), and LSTM network. The DNN model employs a global max pooling layer to manage disparities in sequence lengths. TextCNN uses a 1D convolutional layer. A second layer of ensemble is then applied based on the output of a weighting module to discern the significance of each base learner's contribution to the final classification. This ensemble is not trainable but dynamically

assigns weights. The hierarchical ensemble model's complexity had issues with efficiency and the potential for overfitting [21]. This approach also lacked integrating sentiment cues and highly focuses on semantic cues. There was also an attempt by researchers to collect and analyze a series of posts from social media to identify patterns and indicators of depression. Pre-processing involved BERT embedding and stack-LSTM tagging. An ontology of medical information is built using a single layer Bayesian network [22].

The temporal attention mechanism (TAM)-SenticNet model consists of two principal modules: the neural implementation and the symbolic implementation. This dual structure allows for effective feature extraction and sentiment analysis while incorporating logical reasoning for enhanced interpretability [23]. A hybrid deep learning model (CNN with BiLSTM) for depression detection from social media text analysis attempted to improve the existing methods as they suffer from model representation problems from text with high accuracy, leveraging open X (formerly Twitter) and Reddit datasets from Kaggle. Text embedding using FastText embeddings instead of traditional ones like WordVec. Non-textual features and sentiment cues were not considered [24]. A multi-modal approach for depression detection employing textual CNN, audio CNN, LSTM, and BiLSTM models was attempted. The dataset used was the distress analysis interview corpus - wizard of oz (DAIC-WOZ) depression database with 59 depressed posts and 139 non-depressed posts and an high accuracy of 88% was achieved by the BiLSTM component but achieving only a depression recall of 0.32. The model was also not tested on actual social media data but limited synthetic data and hence generalizability is not established [25].

Based on the above, only a few models integrate sentiment cues explicitly and even in the ones that include sentiment cues. There is scope for future work to leverage approaches like association rule mining [16] which has been explored in this paper. There is also scope to improvise depression recall if the model has to be incorporated in a real time system. This paper proposes a novel SCAB-Net model that addresses these concerns. The other challenge on this domain is the imbalanced nature of datasets. Some of the common datasets used in this domain are listed in Table 1.

Table 1. Common datasets used in related work

Dataset	Features
CLEF 2016 [19]	Consists of title, post content, timestamp, and depressed/nondepressed
Sentiment140 [20]	Username, post content, and sentiment tag
Depressive Tweets Kaggle [22]	Username, timestamp, post text, and depressed/not depressed
eRisk 2018 [22]	Consists of title, post content, timestamp, and depressed/nondepressed
Kaggle Reddit/X [24]	Post content, username, and depressed/nondepressed
DAIC-WOZ [25]	Mix of text, audio, and video samples
eRisk 2017 [26], [27]	Consists of title, post content, timestamp, and depressed/nondepressed

3. METHOD

SCAB-Net is a hybrid neural network architecture that is an ensemble of transformer-based embeddings, sequential modeling, sentiment/emotional context modeling and association rule mining on depressive keywords. It can be broadly referred to as an ensemble of i) linear classifier on top of a neural network that has BERT, BiLSTM, and sentiment aware layer and ii) logistic regression classifier based on association rule mining on top of attention weights from sentiment-aware layer represented in Figure 1. The SCAB-Net model consists of the following components: i) BERT layer, ii) BiLSTM layer, iii) sentiment-aware layer, iv) max pooling and dropout layers, v) classifier layer, vi) association rule mining and logistic regression classifier, and vii) ensemble model.

3.1. Bidirectional encoder representations from transformers layer

The social media text as represented in (1), from the eRisk2017 dataset is passed into the BERT layer (BERT tokenizer), which is represented by (2). This layer creates *input_ids* which are unique identifiers for each token and also *attention_mask* which helps the BERT model differentiate original tokens from padding ones. The output of the tokenizer is passed on to the embedding layer which helps the model understand the words in context. The purpose of this layer is to provide contextualized word embeddings without which, obtaining such embeddings from scratch would be infeasible with limited data.

$$S = [s_1, s_2, \dots, s_L] \text{ where } L \text{ is the length} [100] \quad (1)$$

$$C = BERT(X) = [c_1, c_2, \dots, c_L] \in \mathbb{R}^{(L \times dim)} \quad (2)$$

Where c_t is contextual embedding for token s_t ; d is hidden dimension size (128); $\mathbb{R}$ is set of real numbers; and $\mathbb{R}^{(L \times dim)}$ is output sequence.

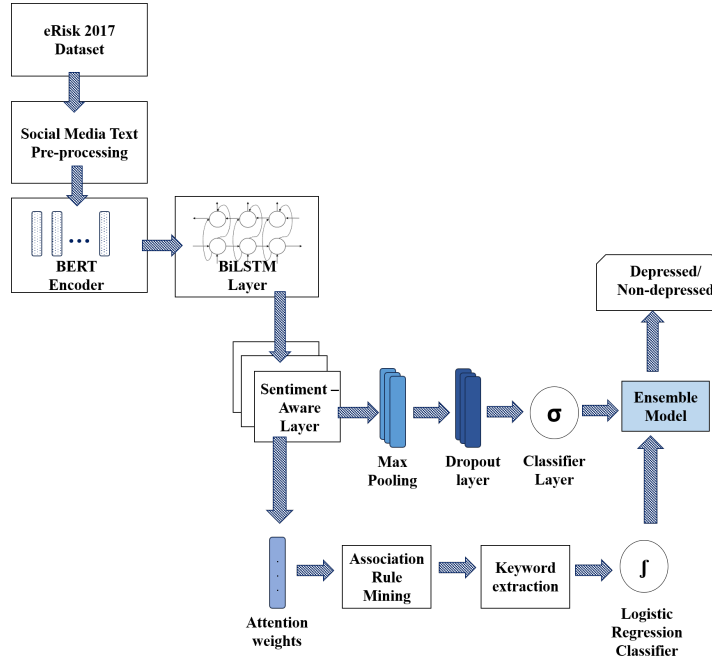


Figure 1. SCAB-Net architecture

3.2. Bidirectional long short-term memory layer

The BiLSTM layer helps the model process the social media text sequence in both forward and backward directions. The forward layer and the backward layers are represented by (3) and (4). The output vector (as in (5)) produced by this layer carries more information about the sequence order. This layer is incorporated on top of the BERT layer to add recurrent inductive bias which is important to capture temporal dependencies in mental state progression. For example: “I used to be jovial” → “Now, I feel hopeless”.

$$\text{Forward hidden state: } \vec{h}_t = LSTM(\vec{h}_t, \vec{h}_{t-1}) \quad (3)$$

$$\text{Backward hidden state: } \overleftarrow{h}_t = LSTM(h_t, h_{t-1}) \quad (4)$$

$$\text{Combining both, } b_t = \{\vec{h}_t, \overleftarrow{h}_t\} \text{ where } b_t = [b_1, b_2, \dots, b_L] \in \mathbb{R}^{(L \times 2h)} \quad (5)$$

Where h is hidden size of one direction of BiLSTM (128).

3.3. Sentiment-aware layer

This layer adds in emotion/sentiment information. It helps to get sentiment scores using valence aware dictionary and sentiment reasoner (VADER) [16] and combines them with the outputs from the BiLSTM layer using (6). This is a lexicon and rule-based sentiment analysis tool specifically designed for social media context which produces four sentiment scores for each text sample – positive, neutral, negative, and compound. Positive, neutral, and negative scores are the proportion of text matching positive, neutral, and negative sentiment lexicons respectively. Compound score is the weighted composite score computed by summing scores of each word in lexicon. The four scores are concatenated into a 4-dimensional sentiment vector for each sample. Based on the sentiment score, attention weights are computed as indicated in (7) which makes the model focus on words that indicate emotion rather than the neutral ones.

$$\text{For each sentiment score } e_t, \text{ apply element wise product/gating, } b'_t = b_t \odot \sigma(W \cdot e_{vader} + \beta) \quad (6)$$

Where W is weight matrix and β is bias.

$$\text{Output of sentiment aware layer: } B' = [b'_1, b'_2, \dots, b'_L] \in \mathbb{R}^{(L \times 2h)} \quad (7)$$

3.4. Max pooling, dropout, and classifier layers

The max pooling layer helps summarize the sequence provided by the previous into a single vector. The dropout layer helps generalize better and helps the model to avoid overfitting. Based on existing work, the datasets in this domain are highly imbalanced and hence regularization is critically important to prevent the neural network from overfitting to majority class patterns. A moderate dropout value of 0.3 was chosen in order to balance regularization strength and model capacity. The output is finally fed into the classifier layer, which uses (8) to output a probability of 0 to 1.

$$\hat{y} = \sigma.(W \cdot u + \beta) \quad (8)$$

Where W is weight matrix and β is bias.

3.5. Association rule mining and logistic regression classifier

The attention weights from the sentiment layer are passed onto this layer. For each depressed case in the training set, extract the words that had the most attention. Top words by highest attention weights are picked for all depressed samples and a transaction matrix is constructed. The occurrences of each word across depressed samples in the training set are marked with a 1 indicating presence of them occurring and 0 in case they do not occur. Standard apriori algorithm is applied for this on top of a transaction matrix to find the frequent item sets. Based on this, association rules are generated. Keywords are extracted from the association rules based on the min_support and min_confidence thresholds set. Based on the keyword features, a logistic regression classifier is employed to predict depressions.

3.6. Ensemble model

The ensemble represented by (9), combines two fundamentally different approaches to leverage their complementary strengths i.e., the output of the classifier layer built on top of sentiment-aware layer (7) and the output of the logistic regression classifier with a decision threshold of 0.5. The ensemble weights are optimized for optimal results. The ensemble is able to have a good accuracy by leveraging the sentiment-aware layer's majority-class performance while also taking into account the association rule mining's minority class-detection. Figure 2 presents a taxonomy diagram that illustrates the complete SCAB-Net pipeline.

$$\hat{y} \times w_a + l \times w_b \quad (9)$$

Where l is output of logistic regression classifier; w_a and w_b are weights.

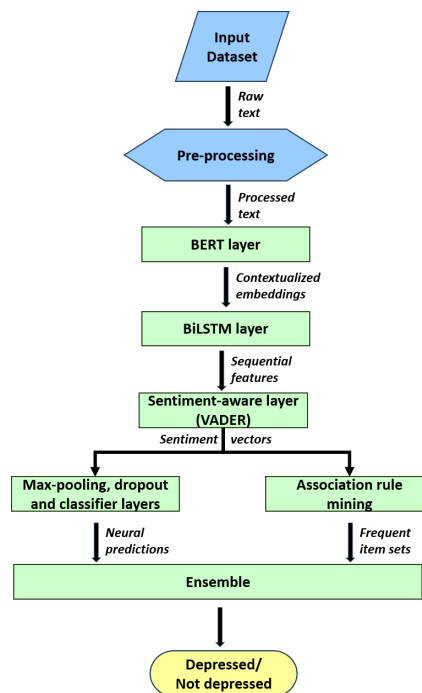


Figure 2. Taxonomy diagram

4. RESULTS AND DISCUSSION

4.1. Experimental setup

Python Torch framework (torch.nn, torch.nn.functional, and torch.utils.data) was used for the implementation leveraging CUDA libraries. Transformers library from Huggingface, BertModel, and BertTokenizer were also used in the implementation. Matplotlib and seaborn were used for plotting and scikit_learn was used for computing various evaluation metrics. Ensemble of both the sentiment-aware neural component and the association rule mining component is used to produce the best possible results leveraging strengths of both approaches. A summary of hyperparameters is presented in Table 2. They were optimized for F1-score on depressed class rather than just “overall accuracy”. For the sentiment-aware layer (including the linear sigmoid classifier), fixed-epoch training strategy was employed and convergence was validated manually after each epoch.

Table 2. Model hyperparameters / configuration

Component	Value
Max. l	128 tokens
BiLSTM layer	128 per direction (256 total)
Sentiment dim	4 (fixed by VADER)
Dropout rate	0.3
Epochs	10
Optimizer	Adam
Learning rate	0.001
Batch size	16
Loss function	BCE
ARM min support	0.005
ARM min confidence	0.3
Ensemble weights	[0.5, 0.5]
Decision threshold	0.5

This was possible because of the following reasons: i) BERT weights were frozen and hence overfitting risk was reduced; ii) dropout layer provides sufficient regularization; and iii) checkpoint saving allows “post-hoc” selection of the model. Bert-base-uncased for used as the BERT tokenizer for tokenizing inputs. This is chosen because uncased is suitable for social media text where we don’t find consistent capitalization. The sequences were thresholded to a max sequence length of 128. Maximum length of 128 balances the efficiency in terms of computation against the information retention i.e, 128 tokens were sufficient to capture majority of the posts while being able to retain reasonable memory requirements for BERT embeddings. The hidden dimensions of 128 in BiLSTM (per direction) provides sufficient capacity to capture sequential dependencies without any overfitting.

A moderate drop out rate of 0.3 is chosen to prevent overfitting of the majority class. Lower values (0.2 and lower) caused overfitting with poor generalization to depressed posts. On the other hand, higher values (0.4 and above) degraded performance on both depressed and non-depressed cases. A learning rate of 0.001 is picked as higher rates led to training instability and divergence specifically for the depressed cases. Lower rates required excessively delayed the convergence in terms of epoch count. Other choices in the proposed architecture include leveraging Adam optimizer and binary cross-entropy (BCE) loss function. Adam optimizer is well suited for sparse gradients which is prevalent in natural language processing. For experimentation purposes, this study tried BCE with focal loss ($\alpha = 0.25$, $\gamma = 2.0$) to cater to the imbalanced dataset but the results weren’t too promising. Hence, standard BCE is leveraged.

In terms of the min_support and min_confidence values for association rule mining, the values were optimized via grid search for classification performance. Some of the values are listed in Table 3. The selected hyperparameters maximized depression F1-score with a feature set of size 25. Similarly ensemble weights were considered based on both theoretical consideration and evaluation. Both the components – sentiment aware neural network and association rule mining exhibit complementary architectures and also error profiles (high precision vs high recall) i.e, the former is good at avoiding false positives whereas the latter is good in capturing true positives. Evaluating these for various ensemble weight configurations showed that [0.5, 0.5] was the most optimal choice. Table 4 lists the sample evaluations. The hardware setup consists of Intel Core i7 processor with 16 GB of RAM and NVIDIA GeForce GTX 1660 Ti GPU with 6 GB VRAM. The dataset used was the eRisk 2017 dataset [26], [27].

Table 3. Validation of min_confidence and min_support parameter choices

Values	Depression F1-score	Depression precision	Depression recall	Features
min_support = 0.001, min_confidence = 0.2	0.223	0.128	0.612	84
min_support = 0.005, min_confidence = 0.3	0.265	0.175	0.550	25
min_support = 0.01, min_confidence = 0.4	0.198	0.221	0.181	12

Table 4. Ensemble weight optimization

Sentiment-aware layer	Association rule mining layer	Depression accuracy	Depression precision	Depression recall	Depression F1-score
0.7	0.3	0.91	0.61	0.18	0.28
0.6	0.4	0.90	0.57	0.21	0.31
0.5	0.5	0.90	0.53	0.24	0.33
0.4	0.6	0.87	0.38	0.35	0.36
0.3	0.7	0.79	0.24	0.48	0.32

4.2. Dataset

The benchmark dataset (eRisk 2017) was used in this research to compare the performance of the SCAB-Net model with existing models. Out of 295,023 entries considered, 30,851 were indicative of depression and 264,172 were normal posts and this data was used to evaluate the performance of the SCAB-Net model and compare it with existing models such as LSTM, Bi-LSTM, CNN, CNN LSTM, CNN Bi-LSTM, CNN with attention, BiLSTM with attention, CNN bidirectional gated recurrent unit (BiGRU). The split was as follows: 70% for training, 30% for the test set.

4.3. Performance comparison and ablation study

The SCAB-Net model outperforms nine baseline models achieving an accuracy of 89.95% and has the best overall performance. The results are presented in Table 5. Figure 3 shows a comparison of the accuracy of various models. The SCAB-Net model excels other models specifically in the minority detection (depressed classes).

Table 5. Performance evaluation

Model	Accuracy	Depression precision	Depression recall	Depression F1-score
Support vector machine (SVM)	0.8790	0.69	0.04	0.07
LSTM	0.8709	0.50	0.17	0.25
BILSTM	0.8699	0.47	0.17	0.25
CNN	0.8665	0.52	0.15	0.23
CNN_LSTM	0.8605	0.38	0.20	0.26
CNN_BILSTM	0.8627	0.42	0.21	0.28
CNN_ATTENTION	0.8747	0.46	0.20	0.27
BILSTM_ATTENTION	0.8731	0.47	0.20	0.29
CNN_BIGRU	0.8641	0.36	0.20	0.26
SCAB-NET	0.8985	0.53	0.24	0.33

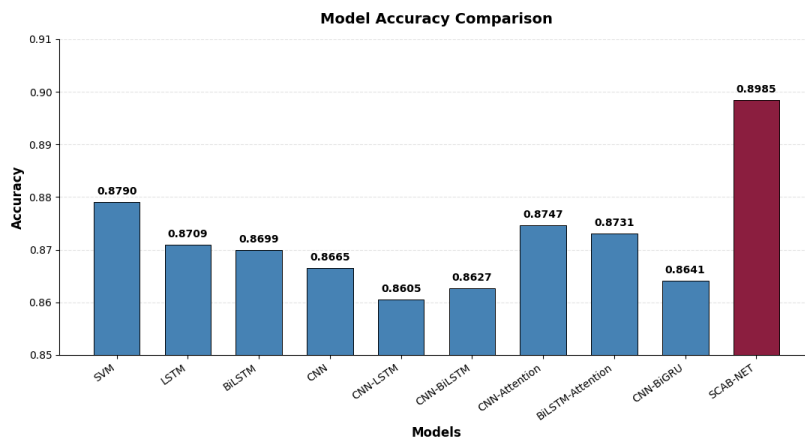


Figure 3. Comparison of accuracy of SCAB-Net against benchmark models

Figure 4 presents the confusion matrix for SCAB-Net, which provides the distribution of correct and incorrect predictions across different classes. To understand the contribution of each component towards overall model performance, an ablation study was performed by starting with the base model (BERT with BiLSTM) and then adding each component one by one. The results are listed in Table 6. Adding a multi-head attention (MHA) layer increased the precision by 27.6% but decreased the overall recall by 26.7% and thus a lower F1-score of 0.20. Hence MHA has been dropped from the architecture. By incorporating sentiment features, the depression recall-precision balance was improved leading to an F1-score of 0.25 (36% depression recall improvement with only a 9% drop in precision). This indicates clearly that the sentiment layer helps the model to identify depression indicators which it previously had missed.

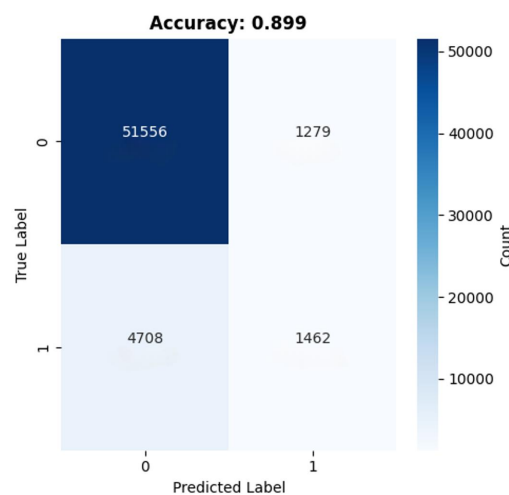


Figure 4. Confusion matrix for SCAB-Net model

Table 6. Ablation study results

Model/component	Accuracy	Depression recall	Depression precision	Depression F1-score
Base model	0.86	0.15	0.58	0.23
Base + MHA	0.86	0.11	0.74	0.20
Base + sentiment-aware layer	0.90	0.15	0.67	0.25
Base + association rule mining	0.68	0.55	0.17	0.27
Base + ensemble	0.90	0.24	0.53	0.33

Furthermore, adding the association rule mining layer indicates a 267% increase in depression recall from 0.15 to 0.55. This clearly tells us that this layer captures certain depression indicators that are not captured by the sentiment-aware neural components. However, there is a drop in precision and overall accuracy. The integration of ensemble with 0.5 weights to both neural and association rule mining models achieves the best F1-score of 0.33 balancing the strengths of both approaches. It increases depression recall from 0.15 to 0.24 (60%) and also ensures decent precision (0.53) and accuracy (0.90).

In order to validate that the performance boost are not random but genuine, paired t-tests were performed and the results are illustrated in Table 7. There are three patterns observed in the t-statistics obtained.

- Minor differences ($|t| < 10$): the difference of 4.93 in base vs base + sentiment aware models indicates not much fundamental changes which is also evident from the similar depression recall values of 0.15.
- Moderate differences ($28 < |t| < 35$): this is seen in the base + ensemble models which indicates that the ensemble produces a good blend of both parent approaches.
- Huge differences ($|t| > 150$): this indicates that orthogonal nature of sentiment-aware and association rule mining components.

All comparisons achieve statistical significance ($p < 0.0001$).

To test the generalizability of the model, the SCAB-Net model was evaluated on a dataset constructed by scraping social media posts from Reddit. The dataset consisted of 100 depressed posts from Reddit. The results show that 80 posts were detected correctly and 20 were missed. Table 8 shows the distribution of error analysis (across sentiments and VADER sentiment scores). Analyzing the missed cases, some of the reasons

to be found were: i) words with censored profanity which caused loss of semantic meaning after tokenization; ii) informal language/internet slangs such as “humble brag”; iii) passive framing vs direct framing: “Praying I never wake up” vs “I want to die”; and iv) grammar errors and non-English patterns.

Table 7. Paired t-test results of various components of SCAB-Net

Model1	Model2	t-statistic
Base model	Base + sentiment-aware	4.93
Base model	Base + association rule mining	-158.44
Base model	Base + ensemble	-28.62
Base + sentiment-aware	Base + association rule mining	-158.74
Base + sentiment-aware	Base + ensemble	-35.24
Base + association rule mining	Base + ensemble	151.03

Table 8. Error analysis

Sentiment	Missed	Detected
Positive	0.102	0.114
Neutral	0.732	0.730
Negative	0.166	0.156
Compound	-0.398	-0.252

4.4. Explainability and ethics

With respect to ethics and explainability, clinicians can see and visualize the exact specific words that led to identification of depressive behavior using attention heatmaps as illustrated in Figure 5. The figure presents the attention visualization for two detected samples, where the intensity of the color shading is directly proportional to the attention weight i.e, darker regions would infer higher attention weights. SCAB-Net model can be used as a means of screening with a human in the loop for reviewing the cases that are flagged. One of the reasons being bias due to VADER sentiment scoring which might have a language bias to English and cultural bias to the Western world. Language is also dynamic in nature leading to a temporal bias. The model does not store or process any personal information such as name and is entirely based on social media text analysis. The dataset used to train the model is a benchmark dataset that does not have any personal information. The dataset used to validate generalization consists of publicly available data from Reddit and raises no ethical concerns.

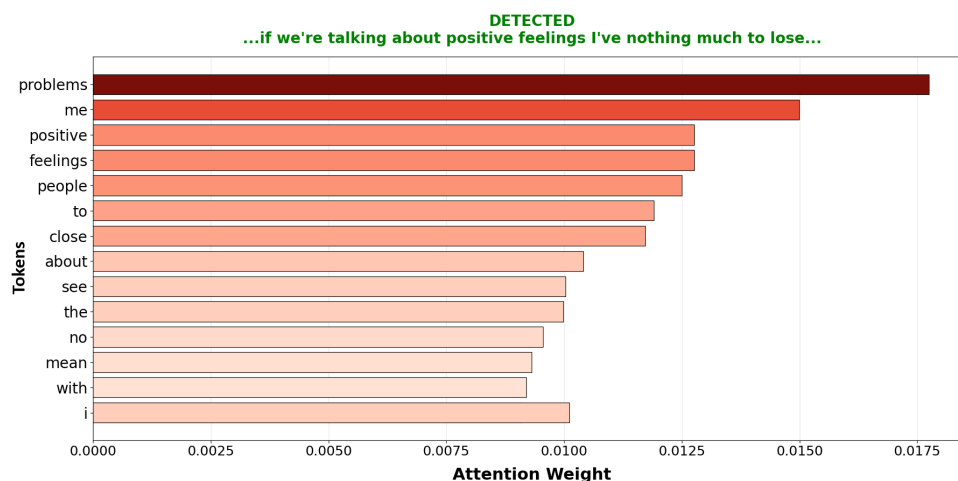


Figure 5. Attention visualization

5. CONCLUSION

In this paper, a novel SCAB-Net was presented to advance sentiment-aware modeling for depression detection from social media posts. This model achieves an accuracy of 0.8985 and has the edge over models such as SVM, LSTM, BiLSTM, CNN, CNN LSTM, CNN Bi-LSTM, CNN with attention, BiLSTM with attention, and CNN BiGRU when evaluated on the benchmark eRisk 2017 dataset. The sentiment-aware

component in isolation improved precision but failed to boost recall values for depressed posts (minority class). Hence, association rule mining component was introduced, which captures orthogonal signals that are otherwise not captured by neural models and an ensemble of these achieved superior balance across metrics. The ensemble of these models boosted the depression recall from 0.15 to 0.24. The work can be potentially extended to other affective computing tasks and used by psychologists to conduct longitudinal studies to track language markers related to depression and other mental health disorders.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Cephas Paul Edward	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Karthikeyan Harimoorthy	✓	✓		✓	✓	✓		✓	✓	✓		✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal Analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project Administration

Fu : Funding Acquisition

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY

The dataset used in this study are from third party (CLEF association). Restrictions apply to the availability of these data, which were used under license for this study. Data is available from eRisk 2017 Text Research Collection at <https://tec.citius.usc.es/ir/code/eRisk2017.html> with the permission of third party.

REFERENCES

- [1] Mya Care, "The mental and physical health struggles of gen z and how to cope," myacare.com, 2023. Accessed: Mar. 01, 2025. [Online]. Available: <https://myacare.com/blog/the-mental-and-physical-health-struggles-of-gen-z-and-how-to-cope>
- [2] N. Dudija and M. D. Apriliansyah, "Psychological well-being among generation Z employees: a literature review," *International Journal of Scientific Research and Management*, vol. 12, no. 08, pp. 79–87, Aug. 2024, doi: 10.18535/ijrm/v12i08.gp01.
- [3] D. Bagaskara, H. Hasanudin, K. Karsim, M. C. I. Kalis, and I. D. I. Kalis, "The influence of artificial intelligence (AI) and information communications technology (ICT) on employee mental health with technostress as mediation in generation z employees in West Kalimantan," *eCo-Buss*, vol. 7, no. 2, pp. 1147–1162, Dec. 2024, doi: 10.32877/eb.v7i2.1781.
- [4] K. Xiang, M.-Y. Wu, Y. Tong, and C. He, "Imprinting and workplace well-being among generation z hotel employees," *Journal of Travel Research*, vol. 64, no. 6, pp. 1393–1413, Jul. 2025, doi: 10.1177/00472875241249396.
- [5] V. A. Gozali, Z. Zamralita, and D. Lie, "Workplace wellbeing and work engagement among generation Z employees in pt x: the role of psychological capital as a moderator," *Biopsikososial: Jurnal Ilmiah Psikologi Fakultas Psikologi Universitas Mercubuana Jakarta*, vol. 8, no. 1, pp. 50–64, Apr. 2024, doi: 10.22441/biopsikososial.v8i1.23750.
- [6] K. M. Keyes, N. T. Kreski, and M. E. Patrick, "Depressive symptoms in adolescence and young adulthood," *JAMA Network Open*, vol. 7, no. 8, Aug. 2024, doi: 10.1001/jamanetworkopen.2024.27748.
- [7] S. Crosier, "Gen Z, social media, and mental health," sph.emory.edu, 2024. Accessed Oct. 02, 2024. [Online]. Available: <https://sph.emory.edu/news/gen-z-social-media-mental-health>
- [8] M. E. Ansell, A. L. F. Jones, D. M. Bayliss, and J. L. Ohan, "'it just makes you feel horrible': a thematic analysis of the stigma experiences of youth with anxiety and depression," *Journal of Child and Family Studies*, vol. 33, no. 7, pp. 2121–2133, Jul. 2024, doi: 10.1007/s10826-024-02877-0.
- [9] E. Samari *et al.*, "Perceived mental illness stigma among family and friends of young people with depression and its role in help-seeking: a qualitative inquiry," *BMC Psychiatry*, vol. 22, no. 1, pp. 107–120, Feb. 2022, doi: 10.1186/s12888-022-03754-0.

- [10] C. Ștefania S. Fantu, "Zetgeneration– smile, empathy, therapy. The opportunities of art and education regarding depression among generation z preadolescents," *Review of Artistic Education*, vol. 27, pp. 188–194, Apr. 2024, doi: 10.35218/rae-2024-0023.
- [11] T. Lyngdoh, D. El-Manstrly, and K. Jeeshia, "Social isolation and social anxiety as drivers of generation Z's willingness to share personal information on social media," *Psychology & Marketing*, vol. 40, no. 1, pp. 5–26, Jan. 2023, doi: 10.1002/mar.21744.
- [12] S. B. F. Paletz *et al.*, "Emotional content and sharing on Facebook: a theory cage match," *Science Advances*, vol. 9, no. 39, Sep. 2023, doi: 10.1126/sciadv.ade9231.
- [13] B. Rimé, "Emotion elicits the social sharing of emotion: theory and empirical review," *Emotion Review*, vol. 1, no. 1, pp. 60–85, Jan. 2009, doi: 10.1177/1754073908097189.
- [14] F. Apriliani and W. Maharani, "Depression detection on social media Twitter using XLNET method," *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 1, pp. 172–180, Feb. 2023, doi: 10.29100/jipi.v8i1.3345.
- [15] S. Alghowinem, T. Gedeon, R. Goecke, J. F. Cohn, and G. Parker, "Interpretation of depression detection models via feature selection methods," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 133–152, 2023, doi: 10.1109/TAFFC.2020.3035535.
- [16] M. Munawar and Y. Yulhendri, "Identification of potential depression in social media posts," *IAES International Journal of Artificial Intelligence*, vol. 14, no. 3, pp. 2096–2101, Jun. 2025, doi: 10.11591/ijai.v14.i3.pp2096-2103.
- [17] W. A. Gadzama, D. Gabi, M. S. Argungu, and H. U. Suru, "The use of machine learning and deep learning models in detecting depression on social media: a systematic literature review," *Personalized Medicine in Psychiatry*, vol. 45–46, Jul. 2024, doi: 10.1016/j.pmp.2024.100125.
- [18] A. G. Ganie and S. Dadvandipour, "Exploring the impact of informal language on sentiment analysis models for social media text using convolutional neural networks," *Multidisciplinária Tudományok*, vol. 13, no. 1, pp. 244–254, 2023, doi: 10.35925/j.multi.2023.1.17.
- [19] J. P. Thekkekkara, S. Yongchareon, and V. Liesaputra, "An attention-based CNN-BiLSTM model for depression detection on social media text," *Expert Systems with Applications*, vol. 249, Sep. 2024, doi: 10.1016/j.eswa.2024.123834.
- [20] S. Dalal, S. Jain, and M. Dave, "Review of advancements in depression detection using social media data," *IEEE Transactions on Computational Social Systems*, vol. 12, no. 1, pp. 77–100, Feb. 2025, doi: 10.1109/TCSS.2024.3448624.
- [21] T. Sheng, W. Cai, J. Huang, and Z. Dong, "Detecting adolescent depression in social media: a hierarchical ensemble learning approach," *Theoretical and Natural Science*, vol. 43, no. 1, pp. 312–325, Jul. 2024, doi: 10.54254/2753-8818/43/20240763.
- [22] W. Zhang, J. Xie, Z. D. Zhang, and X. Liu, "Depression detection using digital traces on social media: a knowledge-aware deep learning approach," *Journal of Management Information Systems*, vol. 41, no. 2, pp. 546–580, Apr. 2024, doi: 10.1080/07421222.2024.2340822.
- [23] R. Dou and X. Kang, "TAM-SenticNet: a neuro-symbolic AI approach for early depression detection via social media analysis," *Computers and Electrical Engineering*, vol. 114, Mar. 2024, doi: 10.1016/j.compeleceng.2023.109071.
- [24] V. Tejaswini, K. S. Babu, and B. Sahoo, "Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 1, pp. 1–20, 2024, doi: 10.1145/3569580.
- [25] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Measurement: Sensors*, vol. 25, Feb. 2023, doi: 10.1016/j.measen.2022.100587.
- [26] D. E. Losada and F. Crestani, "A test collection for research on depression and language use," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction 7th International Conference of the CLEF Association (CLEF 2016)*, Cham, Switzerland: Springer, 2016, pp. 28–39, doi: 10.1007/978-3-319-44564-9_3.
- [27] D. E. Losada, F. Crestani, and J. Parapar, "eRISK 2017: CLEF lab on early risk prediction on the internet: experimental foundations," in *Experimental IR Meets Multilinguality, Multimodality, and Interaction (CLEF 2017)*, Cham, Switzerland: Springer, 2017, pp. 346–360, doi: 10.1007/978-3-319-65813-1_30.

BIOGRAPHIES OF AUTHORS



Cephas Paul Edward    holds a Bachelor of Engineering (B.E.) in Computer Science and Engineering, Master of Engineering (M.E.) in Computer Science and Engineering from Anna University, Chennai. He is currently enrolled as a research scholar in Christ University, Bengaluru, India. He can be contacted at email: cephaspaul.edward@res.christuniversity.in.



Dr. Karthikeyan Harimoorthy    is currently working as assistant professor at Christ University, Bengaluru. He has 13 years of experience in academics. He holds a Bachelor of Engineering (B.E.) in Computer Science and Engineering, Master of Engineering (M.E.) in Computer Science and Engineering, and Ph.D. degree from Anna University, Chennai. His research focuses on using various machine learning algorithms in healthcare industry and cyber security. He can be contacted at email: karthikeyan.h@christuniversity.in.