

# DEVELOPMENT OF A SEMANTIC FEATURE EXTRACTION AND PSO-BASED FEATURE SELECTION MODEL WITH LSTM AND RNN CLASSIFICATION FOR DEPRESSION IN SOCIAL MEDIA TEXT

**Ashwani Kumar<sup>1</sup>**

<sup>1</sup>Research Scholar, Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar 125001, Haryana, India,  
ash.rohila1985@gmail.com

**Sunita Beniwal<sup>2</sup>**

<sup>2</sup>Assistant Professor Department of Computer Science and Engineering, Guru Jambheshwar University of Science and Technology, Hisar 125001, Haryana, India,  
2Corresponding author: sunitabeniwalcse@gmail.com

**Abstract** - Social media text is short, informal, and noisy, and these properties make the feature engineering stage difficult in depression-related text analysis. This study focuses on semantic feature extraction and feature selection. A GloVe feature extraction and PSO feature selection framework is developed in which Global Vectors for Word Representation (GloVe) converts preprocessed tweets into dense semantic vectors, and Particle Swarm Optimization (PSO) selects the most informative dimensions from that vector space. The optimized feature set is validated using two comparative configurations, namely GloVe-PSO with LSTM and GloVe-PSO with RNN, to verify whether the selected features improve class separability. Kaggle dataset used in this study contains more than 20,000 posts; after cleaning and filtering, 9,075 records were retained for embedding and validation. The original 100-dimensional GloVe representation was reduced to 43 selected semantic features. The optimized feature set achieved strong downstream validation performance, with RNN giving the best evaluation: 89.37% accuracy, 90.66% precision, 96.79% recall, and 93.62% F1-score, while the feature extraction-selection stage required 30 seconds. The GloVe-PSO with RNN configuration outperformed GloVe-PSO with LSTM in the comparative validation. The results indicate that optimizing semantic embedding dimensions can reduce redundancy and improve the practical usefulness of features for depression-related social media analysis.

**Keywords** - Depression-related text analysis, feature extraction, feature selection, Particle Swarm Optimization, social media text.

## 1. Introduction

Social media has proven to be a helpful tool for the study of depression-related language, as people tend to share their sadness, hopelessness, fatigue and social withdrawal in their normal posts. A recent systematic review and meta-analysis revealed that textual content, language style, social media activity and temporal features all have an impact on depression prediction, but also highlighted that the validity of the models relies on the representation and evaluation of these

features [1]. In a 2025 survey, it was found that feature extraction in social media depression studies is largely based on models such as FastText, Word2Vec, BERT, and their variants, and that post-level analysis remains a challenge due to the brevity and variability of expressions [2]. The feature stage is not just a minor preprocessing step; it's where noisy text becomes measurable evidence. Hameed et al. demonstrated that performance varies significantly with the feature set and the type of classifier used, and that GloVe features sometimes outperform purely lexical features in their comparisons [3]. This observation is in line with the present study, which does not consider feature extraction and feature selection as a supporting detail, but rather as the main research problem.

The proposed work uses GloVe, which models semantic relationships between words using global word co-occurrence statistics [4], and applies PSO, a population-based search method originally developed for continuous optimization [5], to select a compact subset of useful embedding dimensions. In this study, GloVe is used to obtain the semantic representation, and PSO searches that representation to remove redundant dimensions. GloVe-PSO with LSTM and GloVe-PSO with RNN are then used as evaluators of feature quality to identify depression.

The primary goal is to create a smaller and more meaningful feature space for social media text related to depression. This is also a way of not overstating the work as a clinical diagnostic system. The result of the framework should be interpreted as an optimized feature representation which can be used for the subsequent classification and analysis.

## Objectives and Contributions

The specific contributions of the paper are as follows:

- To extract dense semantic features from cleaned depression-related social media posts using GloVe embeddings;
- To reduce redundant and weak semantic dimensions using PSO while preserving features useful for downstream classification;
- To validate the selected feature subset using GloVe-PSO with LSTM and GloVe-PSO with RNN through standard metrics, including accuracy, precision, recall and F1-score.

## 2. Literature Review

Recent work on detection of depression have used machine learning, deep learning and Natural language processing techniques for analysis of text on social media. The main challenge is to represent textual information effectively reducing feature redundancy and computational complexity.

### Depression Detection Using Machine Learning and Deep Learning

Deep learning architectures have been used for identification of depressive patterns in social media data. [6-8] Sugiyanto et al. demonstrated the importance of feature preservation when the input representation is sensitive to information loss by employing a pooling-less CNN and facial Action Unit intensity features. [6]. Suganthi and Geetha proposed an optimized machine learning model for prediction of postpartum depression from social media text and feature-level linguistic information [7]. Thekkekara et al. proposed an attention-based CNN-BiLSTM model that learned

contextual patterns from social media text [8]. Although these studies address different modalities, their emphasis is on feature quality in improving classification performance.

Transformer-based models have also been used recently for social media studies. Kerasiotis et al. used transformer representations and auxiliary features and data augmentation for depression detection [9]. Qasim et al. investigated transformer-based sentence embeddings for depression severity classification [10] and Hussain et al. extended transformer-based methods for multi-level depression severity detection [11]. Alambo et al. [12] showed that contextualized language models based on BERT can effectively capture semantic and psychological information from social media text. These studies demonstrate the usefulness of contextual representation, but their high-dimensional embeddings are not necessarily compact.

### **Feature Selection and Optimization methods:**

There are several challenges associated with high-dimensional feature space and researchers have investigated feature selection and optimization approaches for addressing these issues. Baydili et al. used feature selection as part of a deep learning pipeline for depression and suicidal tendency detection in social media data [13]. Effendi et al. applied PSO and Random Forest for student depression prediction, showing the usefulness of PSO in selecting mental-health-related features [14]. In the field of EEG-based depression recognition, Wang et al. used PSO to enhance feature optimization for classifiers, and they demonstrated that PSO can also be used to enhance feature optimization for classifier performance enhancement. [15]. These studies shows that selection of informative features can reduce dimensionality, improve computation efficiency and classification accuracy.

### **Swarm Intelligence and Metaheuristic optimization**

Swarm intelligence and metaheuristic algorithms are used for feature refinement and optimization in depression analysis. Atta et al. conducted a study on swarm based neural networks for automated depression detection [16]. Geetha and Kumar introduced a model using the cockroach swarm optimization technique to extract optimal keywords for predicting depression from Twitter [17]. Chitrakala and Senthamizhselvi applied Gannet Optimization to optimize code-mixed features for multi-emotion depression classification [18]. These studies show the effectiveness of nature-inspired algorithms in improving predictive performance by identifying relevant features.

Explainable AI and graph-based learning have been combined with optimization approaches. Thamrin and Chen enhanced explainability in depression detection from social media using feature engineering and LLMs [19]. Mao and Han proposed a textGCN framework that incorporated emotional relationships into graph-based textual representations [20]. Arshad et al. created a multi-platform method that relies on careful preprocessing and contextual analysis of Twitter, Reddit, and Facebook data [21]. These Approaches provide promising results by optimizing keywords, handcrafting attributes and classifier parameters rather than refining dense semantic embedding spaces directly. In Table 1, the recent studies related to depression analysis are summarized.

**Table 1. Summary of recent studies related to feature extraction and optimization**

**DEVELOPMENT OF A SEMANTIC FEATURE EXTRACTION AND PSO-BASED FEATURE SELECTION MODEL WITH LSTM AND RNN CLASSIFICATION FOR DEPRESSION IN SOCIAL MEDIA TEXT**

| Ref. | Method                                  | Feature/optimization focus                                 | Relevance to this study                                                             |
|------|-----------------------------------------|------------------------------------------------------------|-------------------------------------------------------------------------------------|
| [6]  | CNN pooling-less model                  | Preservation of facial Action Unit features                | Shows that retaining informative features can improve depression-related modelling. |
| [7]  | Optimized MLP for postpartum depression | LIWC and LSA-based social media features                   | Uses optimized learning with text-derived features.                                 |
| [9]  | Transformer + auxiliary features        | Transformer embeddings and metadata/linguistic markers     | Highlights the value of rich representations but not embedding selection.           |
| [8]  | Attention CNN-BiLSTM                    | Context-aware social media text features                   | Captures sequence patterns but uses a larger deep architecture.                     |
| [10] | Sentence transformer models             | Contextual sentence embeddings                             | Emphasizes semantic embeddings for severity classification.                         |
| [11] | Deep transformers                       | Multi-level depression severity features                   | Shows strong contextual modelling but high feature complexity.                      |
| [13] | Deep learning with feature selection    | Selected features from social media depression data        | Directly supports the need for feature selection.                                   |
| [14] | RF + PSO                                | PSO-based feature selection                                | Demonstrates PSO for mental-health feature optimization.                            |
| [15] | PSO-SVM                                 | PSO-based optimization for depression recognition          | Shows PSO utility in a different modality.                                          |
| [16] | Swarm-based neural networks             | Metaheuristic optimization                                 | Supports swarm intelligence for automated depression detection.                     |
| [17] | COPLELM                                 | Optimal keyword extraction by cockroach swarm optimization | Relevant IJETT study using optimization in social media depression text.            |
| [18] | GO-DEPRESSION                           | FastText features refined by Gannet Optimization           | Recent optimization-based feature refinement for code-mixed text.                   |
| [19] | Feature engineering + LLMs              | Engineered emotion and text features                       | Links feature engineering with explainability.                                      |
| [21] | Multi-platform contextual analysis      | Preprocessing and contextual feature analysis              | Shows the need for robust feature preparation across platforms.                     |

## Research Gap

Despite significant progress in depression detection using deep learning, transformer architectures and metaheuristic optimization techniques, several challenges still exist. Existing research either applies optimization techniques on handcrafted features or employs high-dimensional semantic embeddings without feature optimization. Hence leading to redundant information in feature space, increasing computational cost and decreasing model interpretability. Therefore, there is a need for a compact and computationally efficient framework that can directly optimize semantic embeddings while preserving contextual information. To tackle these issues, the present study integrates GloVe-based semantic feature extraction with PSO-based feature subset selection to generate an optimized feature set for depression detection.

### **3. Materials and Methods**

The main focus of the proposed methodology is on the construction of an optimized semantic feature representation for social media text related to depression. The conventional depression detection frameworks emphasize classifier development mainly, but the present study focuses on feature extraction, dimensionality reduction, and feature quality evaluation. The workflow of this study has five major stages: (i) data acquisition, (ii) text normalization and preprocessing, (iii) semantic representation learning using GloVe embeddings, (iv) dimensionality optimization through Particle Swarm Optimization (PSO), and (v) feature quality validation using deep learning classifiers.

#### **Dataset Description**

The Mental Health Social Media dataset from Kaggle [22] was used for the experiments. This dataset includes over 20,000 English tweets associated with emotional well-being, depression, stress, and anxiety and their corresponding text, metadata and class labels. The original fields are: post creation time, post identifier, user identifier, post text, follower count, friend count, favorites, statuses, retweets, and label. No user identifiers were used as modeling features. There are three categories of the label field: positive, negative or depression/distress-related, and neutral. As this work is centered on feature extraction and selection, the labels are used to assess the improvement in downstream class separation resulting from the selected features. To improve data reliability, duplicate records, incomplete entries, and extremely short posts were removed. After this process, the final retained corpus used for embedding contained 9075 records. The dataset was utilized for computational analysis of textual patterns.

#### **Table 2 Sample Records from Dataset [20]**

**DEVELOPMENT OF A SEMANTIC FEATURE EXTRACTION AND PSO-BASED FEATURE SELECTION MODEL  
WITH LSTM AND RNN CLASSIFICATION FOR DEPRESSION IN SOCIAL MEDIA TEXT**

| Sr No. | Post ID  | Post Created                   | Post Text                                                                                                                                     | User Id  | Followers | Friends | Favorites | Stat uses | Retweets | Label |
|--------|----------|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|----------|-----------|---------|-----------|-----------|----------|-------|
| 0      | 6.38E+17 | Sun Aug 30 07:48:37 +0000 2015 | It's just over 2 years since I was diagnosed with #anxiety and #depression. Today, I'm taking a moment to reflect on how far I've come since. | 1.01E+09 | 84        | 211     | 251       | 837       | 0        | 1     |
| 1      | 6.38E+17 | Sun Aug 30 07:31:33 +0000 2015 | It's Sunday, and I need a break, so I plan to spend as little time as possible on the #A14...                                                 | 1.01E+09 | 84        | 211     | 251       | 837       | 1        | 1     |
| 2      | 6.38E+17 | Sat Aug 29 22:11:07 +0000 2015 | I am awake but tired. I need to sleep, but my brain has other ideas...                                                                        | 1.01E+09 | 84        | 211     | 251       | 837       | 0        | 1     |
| 3      | 6.38E+17 | Sat Aug 29 18:40:49 +0000 2015 | RT @SewHQ: #Retro bears make perfect gifts and are great for beginners too! Get stitching with Octob...                                       | 1.01E+09 | 84        | 211     | 251       | 837       | 2        | 1     |
| 4      | 6.38E+17 | Sat Aug 29 18:40:26 +0000 2015 | It's hard to say whether packing lists are making life easier or just reinforcing how much is still needed...                                 | 1.01E+09 | 84        | 211     | 251       | 837       | 1        | 1     |

Table 3 shows the sample cleaned social media posts after the noisy elements like URLs, mentions, hashtags, and special characters are removed before the feature extraction process.

**Table 3. Sample cleaned posts after removing unwanted text**

| Index | Post_text                                        | Label |
|-------|--------------------------------------------------|-------|
| 0     | It is just over 2 years since I was diagnosed... | 1     |

| Ind<br>ex | Post_text                                             | Lab<br>el |
|-----------|-------------------------------------------------------|-----------|
| 1         | It is Sunday; I need a break, so I am planning...     | 1         |
| 2         | Awake but tired. I need to sleep but my brain...      | 1         |
| 3         | RT @SewHQ: Retro bears make perfect gifts and...      | 1         |
| 4         | It is hard to say whether packing lists are making... | 1         |

### Text Normalization and Corpus Preparation

The preprocessing step converted noisy posts into a cleaner standardized linguistic form before feature extraction, which involved removing URLs, hashtags, numbers, user mentions, and special characters. The text was then tokenized, stop words were removed and lemmatization was applied to reduce different word forms to a common base. The examples of stop-word removal, post-lemmatization, and filtration are shown in Fig. 2, Fig. 3, and Fig. 4, respectively.

Filtering was applied before vectorization so that GloVe embeddings were generated only for cleaned posts retained in the final corpus. The resulting corpus consisted of normalized textual instances suitable semantic embedding generation.

```

0 [year, since, diagnosed, Today, Im, taking, mo.
1 [Sunday, need, break, Im, planning, spend, lit.
2 [Awake, tired, need, sleep, brain, ide
3 [RT, bear, make, perfect, gift, great, beginne.
4 [hard, say, whether, packing, list, making, li.

```

Fig. 2 Stop keyword removal from post\_text data

```

post_text ... filtered_text_cleaned
0 It's just over 2 years since I was diagnosed w... [year, since, diagnosed, Today, Im, taking, mo...
1 It's Sunday, I need a break, so I'm planning t... [Sunday, need, break, Im, planning, spend, lit...
2 Awake but tired. I need to sleep but my brain ... [Awake, tired, need, sleep, brain, idea]
3 RT @SewHQ: #Retro bears make perfect gifts and... [RT, bear, make, perfect, gift, great, beginne...
4 It's hard to say whether packing lists are mak... [hard, say, whether, packing, list, making, li...

```

Fig. 3 Post-lemmatization data

```

0 [years, since, diagnosed, Today, Im, taking, m...
1 [Sunday, need, break, Im, planning, spend, lit...
2 [Awake, tired, need, sleep, brain, ideas]
3 [RT, bears, make, perfect, gifts, great, begin...
4 [hard, say, whether, packing, lists, making, l...

```

Fig. 4 Filtration text data values

### Semantic Representation Learning

The cleaned text was converted to dense semantic vectors using Global Vectors for Word Representations (GloVe) instead of using handcrafted features. The method learns word

representations from global co-occurrence information and projects textual information into continuous vector space thus words that are semantically related occupy closer locations. This property is useful for text related to depression as the same emotion can be expressed in different surface forms.

Each social media post was transformed into a fixed-length vector representation. The resulting text was represented as a feature matrix of 9075 rows by 100 columns that encoded contextual and semantic relationships among words related to depression. This matrix was used as the input to the feature selection stage. The extracted GloVe feature matrix is displayed in a snapshot in Fig. 5. The goal of the GloVe stage was not to classify the posts directly, but to give a dense semantic representation that PSO can search.

```
df_final['glove_vector']
0      [0.15352055268954032, -0.00044607054268088294,...
1      [0.1188597636688168, -0.001597337679760087, -0...
2      [0.11477789144891386, -0.024342390081773583, 0...
3      [0.12364046035095229, 0.0003717089960471206, -...
4      [0.1603133392240028, 0.029701109720080828, -0...
...
9070   [0.1402984260425481, 0.041455825456892546, -0...
9071   [0.12207714616505534, 0.02829853253755367, -0...
9072   [0.1304357309808682, 0.03227481360706251, -0...
9073   [0.18298196405080627, -0.013215913056311283, -...
9074   [0.1725826487537812, 0.05508019431409511, -0.1...
```

Fig. 5 Extracted GloVe feature snapshot

The GloVe objective can be described as minimizing the weighted difference between the inner product of word vectors and the logarithm of their co-occurrence counts. Simplified form as shown in eq (i),

$$J = \sum_{i,j=1}^V f(X_{ij})(w_i^T w_j + b_i + b_j - \log X_{ij})^2 \quad (i)$$

where  $X_{ij}$  represents word co-occurrence. The graphical distribution of the extracted semantic features is presented in Fig. 6.

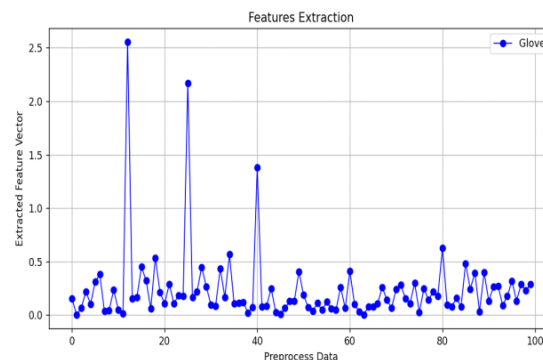


Fig. 6 Graphical representation of extracted features

### Semantic dimension optimization using PSO

Semantic embeddings provide rich contextual information, many dimensions may contain redundant or less informative information. So, PSO was then applied as a wrapper method over the 100 GloVe dimensions to identify most discriminative features.

Each particle represented a candidate feature subset of embedding dimension, where each dimension could be selected or excluded. The swarm iteratively explores search space by updating

candidate solutions according to personal best and global best positions. The optimization objective is to maximize feature relevance and minimize redundancy. In the present work, the validation F1-score of the classifier was used to guide the search.

The particle update can be summarized as

$$\begin{aligned} v\_i(t+1) &= wv\_i(t) + c1r1[p\_i(t) - x\_i(t)] + c2r2[g(t) - x\_i(t)] \\ x\_i(t+1) &= x\_i(t) + v\_i(t) \end{aligned} \quad (ii)$$

Here in eq (ii),  $x\_i$  is a particle position,  $v\_i$  is velocity,  $p\_i$  is the personal best position and  $g$  is the global best position. The fitness function was defined as  $\text{Fitness} = 1 - F1$ . A lower value therefore indicates a better selected feature subset. The optimized collection of selected features produced by PSO is shown in Fig. 7.

```
x_testing
array([[ 0.13121592,  0.04503582, -0.41563573, ...,  0.14746812,
        -0.18661936,  0.32207968]],

       [[ 0.2487037 , -0.16484898, -0.45111625, ...,  0.187405 ,
        -0.18246458,  0.10661284]],

       [[ 0.16446996, -0.04193739, -0.40587396, ...,  0.14544598,
        -0.26191386,  0.25099972]]]
```

**Fig. 7 Optimized collection of selected features**

Table 4 lists the PSO hyperparameters used to guide the feature selection process over the GloVe embedding dimensions.

**Table 4. PSO algorithm hyperparameters**

| Parameter     | Value |
|---------------|-------|
| Iteration     | 100   |
| No. of swarms | 5     |
| Vmax          | 4     |
| Vmin          | -4    |
| C1            | 0.5   |
| C2            | 0.5   |
| R1            | 0.5   |
| R2            | 0.5   |
| Wg            | 0.25  |

The final selected subset contained 43 features from the original 100-dimensional GloVe representation. The reduction indicates that a large part of the dense embedding space was either redundant or less useful for the downstream evaluation task. The selected optimized feature distribution is shown in Fig. 8.

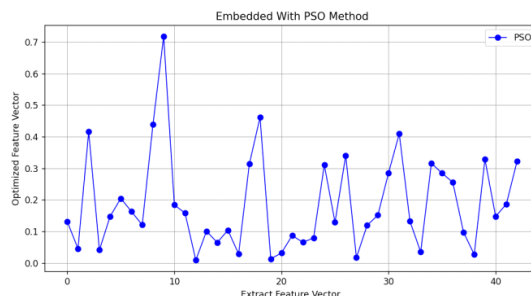


Fig. 8 Graphical representation of selected optimized features

### Feature Quality Validation

To assess the effectiveness of optimized semantic representation, two commonly used sequence learning models, namely Long-short-term-memory (LSTM) and Recurrent Neural Network (RNN) were used. The optimized feature subset was supplied to both classifiers under identical experimental settings. The evaluation metrics used were accuracy, precision, recall and F1-score.

### Experimental Evaluation Process

The dataset was partitioned into training and test subsets using an 80:20 ratio. Feature optimization was performed on training subset to avoid leakage of information. The optimized feature representation was tested on unseen test samples.

Computational indicators like mean, standard deviation, error rate, and the execution time are also analyzed to assess the efficiency and stability of the proposed work. The overall methodology hence emphasizes on semantic representation quality and feature optimization rather than improvements in classifier

## 4. Results and Discussion

This section evaluates the effectiveness of the proposed GloVe-PSO semantic feature optimization framework. The analysis focusses on feature stability, dimensionality reduction and discriminative capability of optimized feature subset. The selected features were further validated using LSTM and RNN to test whether reduced semantic representations retained useful information.

### Experimental Setup and Performance Metrics

The design and implementation of the GloVe-PSO feature selection framework and the comparative LSTM and RNN validation configurations were executed on a windows 10 (64-bit) platform equipped with INTEL Core™ i7 CPU@3GHz processor having 8GB RAM, and 500 TB HDD executing and carried out in Python 3.12. Various statistical and classification metrics were employed to assess the performance. Mean and standard deviation were used to assess the stability and consistency of selected features. Execution time was used to measure computational efficiency. Accuracy, precision, recall and f1-score were used to evaluate the validation models.

Mean ( ): An indicator of central tendency, mean is used to assess the average behavior of all samples. It can be shown as follows in eq (iii).

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (iii)$$

Where,  $\mu$  = dataset mean,  $x_i$  = individual value, and  $n$  = no. of observations.

Standard Deviation (SD) ( $\sigma$ ): The dataset's degree of variation or distribution is measured by Standard Deviation. Reliable results are indicated by a lower value of SD. It can be represented as in eq (iv).

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (\text{iv})$$

Where,  $\mu$  = dataset mean,  $x_i$  = individual value, and  $n$  = no. of observations.

Time Consumption (T): The duration of time required to complete feature extraction and model training and testing is measured by T. It displays the method's computational effectiveness. It can be shown as in eq (v).

$$T = t_{\text{end}} - t_{\text{start}} \quad (\text{v})$$

Where,  $t_{\text{start}}$  = time stamp at the beginning of execution and  $t_{\text{end}}$  = time stamp at the end of execution.

Accuracy: It means the overall exactness of the classification (LSTM and RNN) model. It is defined by the ratio of accurately predicted samples to the total no. of data samples.

$$Acc = \frac{Tp + Tn}{Tp + Fp + Fn + Tn} \quad (\text{vi})$$

Here TP: accurately predicted positive data samples, TN: accurately predicted negative data samples, FP: negative data samples inaccurately predicted as positive, and FN: positive data samples inaccurately predicted as negative.

Precision: It calculates the exactness of positive predictions.

$$pre = \frac{Tp}{Tp + Fp} \quad (\text{vii})$$

Recall: It is also called as Sensitivity (SN) and it calculates the ability of the model to accurately identify actual positive samples.

$$Rec = \frac{Tp}{Tp + Fn} \quad (\text{viii})$$

F1-score: It is the harmonic mean of precision and recall. It shows a balanced calculation of model performance, normally when the database is imbalanced.

$$f1 - score = \frac{2 * pre * rec}{pre + rec} \quad (\text{ix})$$

### Stability of the Feature Representation

The mean and standard deviation values were calculated across multiple runs to investigate whether the optimized feature representation behaved consistently. Figures 9 and 10 represent the mean and standard deviation trends, indicating that selected semantic features preserve contextual information and low values of SD indicate PSO based selection produces stable feature subsets with minimum variation across repeated executions. The execution time required for feature extraction and optimization process is shown in Fig. 11. As shown, the feature extraction-selection pipeline was completed in 30 seconds. This execution time suggests that the GloVe-PSO step is feasible for the dataset size used in this experiment and does not require the computational burden associated with very large transformer pipelines.

These measures describe the stability of the feature extraction and selection process, whereas classification metrics assess the discriminative usefulness of the selected features.

# DEVELOPMENT OF A SEMANTIC FEATURE EXTRACTION AND PSO-BASED FEATURE SELECTION MODEL WITH LSTM AND RNN CLASSIFICATION FOR DEPRESSION IN SOCIAL MEDIA TEXT

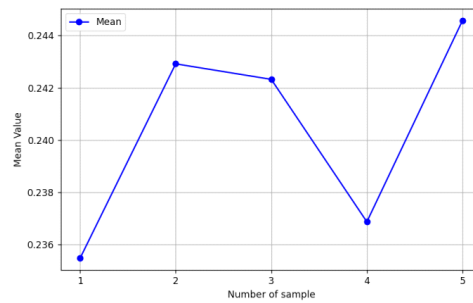


Fig. 9 Mean calculation based on the GloVe and PSO feature representation

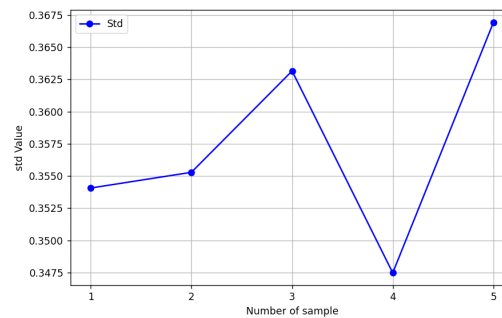


Fig. 10 Standard deviation based on the GloVe and PSO feature representation



Fig. 11 Time taken based on the GloVe and PSO feature selection framework

## Downstream Validation of GloVe-PSO with LSTM and RNN Model

To assess the reliability of selected semantic dimensions, the optimized feature subset was fed to an RNN and an LSTM. The primary objective was not to compare the classifiers but to determine the discriminative capability of selected features. Table 5 shows the performance of both classifiers. The results show that both models achieved comparable performance when trained using optimized features. RNN achieved the highest overall performance. The result indicates that the selected semantic features are informative and independent of a specific classifier.

**Table 5. Downstream validation using GloVe-PSO with RNN and GloVe-PSO with LSTM**

| Metric (%) | GloVe-PSO with RNN | GloVe-PSO with LSTM |
|------------|--------------------|---------------------|
| Accuracy   | 89.37              | 88.48               |

| Metric (%) | GloVe-PSO with RNN | GloVe-PSO with LSTM |
|------------|--------------------|---------------------|
| Precision  | 90.66              | 91.15               |
| Recall     | 96.79              | 94.95               |
| F1-score   | 93.62              | 93.01               |
| Error rate | 10.63              | 11.52               |

### Comparative analysis with existing Optimization-Based Approaches

The comparison in Table 6 is included to position the feature selection framework against related optimization-based approaches. The results reveal that the proposed GloVe-PSO framework consistently produces better precision, recall and f1-score as compared to traditional optimization-based methods. The proposed framework benefits from integrating semantic embedding and evolutionary feature selection, enabling it to retain informative dimensions and remove redundant information. The proposed framework optimizes semantic embedding dimensions directly, unlike existing approaches, hence enhancing compactness of features and quality of representation.

**Table 6. Indicative comparison with optimization-based depression-analysis methods**

| Methods             | Acc.  | Pre.  | Rec.  | F1    |
|---------------------|-------|-------|-------|-------|
| BiLSTM [23]         | 72.85 | 73.88 | 72.37 | 73.12 |
| Dragonfly+NN [16]   | N/R   | 74    | 80    | 76    |
| Firefly+NN [16]     | N/R   | 85    | 88    | 86    |
| MFO+N [16]          | N/R   | 72    | 79    | 80    |
| LSTM [24]           | 69.92 | 69.62 | 68.34 | 68.63 |
| GloVe-PSO with RNN  | 89.37 | 90.66 | 96.79 | 93.62 |
| GloVe-PSO with LSTM | 88.48 | 91.15 | 94.95 | 93.01 |

### Effectiveness of proposed approach

The primary objective of this study was to decrease the dimensionality without removing contextual information. The original GloVe representation had 100 embedding dimensions. After optimization PSO selected only 43 informative dimensions, hence reducing by approximately 57%. While dense embeddings are beneficial for representing semantic relationships, they also have the drawback that some dimensions may be weak, noisy or redundant for a particular task. To solve this problem, PSO searches for a selected subset of embedding dimensions. The fitness function is based on the validation F1-score obtained through the GloVe-PSO with LSTM and GloVe-PSO with RNN configurations, which means that the search is guided by downstream class separability rather than by dimensionality reduction alone.

This design provides a clear role for the method: GloVe provides semantic coverage, PSO eliminates redundant dimensions, and the GloVe-PSO with LSTM and GloVe-PSO with RNN

configurations verify whether the selected features contain useful class information. The method is thus at the feature engineering level, as stated in the aim of this paper.

### **Discussion**

The results confirm that a smaller subset of semantic features can be more useful than the full embedding matrix. The representation was reduced from 100 to 43 dimensions, and validation with GloVe-PSO with LSTM and GloVe-PSO with RNN showed no performance degradation. This implies that some dimensions in the original embedding were not useful for this dataset and that PSO found a more useful subset of dimensions.

The study also reveals the need to report feature selection separately from the design of the downstream validators. Since GloVe-PSO with LSTM and GloVe-PSO with RNN are both evaluated using the same selected feature subset, the improvement is more easily attributable to the feature set than to a single classifier.

### **5. Conclusion and Future Scope**

This paper presented a GloVe-PSO system for optimizing semantic features of social media text related to depression. The proposed approach integrated GloVe based semantic representation with PSO to create a compact and informative feature space from high-dimensional textual embeddings. GloVe was used to embed cleaned posts into a dense semantic vector space, and PSO was used to choose the most informative dimensions from this embedding space. This study focuses on feature extraction and feature quality assessment rather than on classifier design.

The results demonstrated that PSO reduced original GloVe embedding space from 100 dimensions to 43 highly informative semantic dimensions. Statistical analysis proved the stability and computational efficiency of optimized feature representation. LSTM and RNN were used for validation of selected features. The best downstream validation performance was obtained by GloVe-PSO with RNN, which reported 89.37% accuracy, 90.66% precision, 96.79% recall, 93.62% F1-score, and 30 seconds execution time while GloVe-PSO with LSTM also produced comparable results. The consistent performance across both classifiers confirms that meaningful information required for depression classification is preserved. These findings imply that optimized semantic representations can give a more compact and computationally efficient feature space. There are some restrictions, however. The data is not from clinical interviews, but from social media labels; the results should not be interpreted as diagnosis. External comparison is restricted due to variations in data sets and protocols.

In future, the proposed framework can be assessed on multilingual and cross-domain social media datasets to assess its generalization capability. Other metaheuristic optimization algorithms like genetic algorithm, grey wolf optimization, whale optimization, firefly optimization and hybrid optimization strategies can be evaluated. Explainable AI techniques may be used to identify linguistic and semantic features preserved within the optimized feature subset.

### **Conflict of Interest**

The authors declare that they have no conflicts of interest.

### **Author Contributions**

The first author contributed to literature review, methodology design, dataset preparation, implementation, evaluation, drafting and visualization. The second author supervised the work, reviewed the methodology and supported manuscript correction and project guidance.

## References

- D. Phiri, F. Makowa, V. L. Amelia, Y. V. A. Phiri, L. P. Dlamini, and M. H. Chung, “Text-Based Depression Prediction on Social Media Using Machine Learning: Systematic Review and Meta-Analysis,” *J. Med. Internet Res.*, vol. 27, p. e59002, 2025. doi:10.2196/59002.
- P. Ta, N. Tran, H. Nguyen, and H. D. Nguyen, “Detecting Signs of Depression on Social Media: A Machine Learning Analysis and Evaluation,” *Sustainable Futures*, 2025. doi:10.1016/j.sfsr.2025.100827.
- S. Hameed, M. Nauman, N. Akhtar, M. A. B. Fayyaz, and R. Nawaz, “Explainable AI-Driven Depression Detection from Social Media Using Natural Language Processing and Black Box Machine Learning Models,” *Front. Artif. Intell.*, vol. 8, article 1627078, 2025. doi:10.3389/frai.2025.1627078.
- J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation,” in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.
- J. Kennedy and R. Eberhart, “Particle Swarm Optimization,” in *Proc. IEEE Int. Conf. Neural Networks (ICNN’95)*, 1995, pp. 1942–1948.
- S. Sugiyanto, I. K. E. Purnama, E. M. Yuniarno, W. Anggraeni, and M. H. Purnomo, “Depression Classification Based on Facial Action Unit Intensity Features Using CNN-Poolingless Framework,” *Int. J. Intell. Eng. Syst.*, vol. 17, no. 5, pp. 172–187, 2024. doi:10.22266/ijies2024.1031.15.
- D. Suganthi and A. Geetha, “Predicting Postpartum Depression with Aid of Social Media Texts Using Optimized Machine Learning Model,” *Int. J. Intell. Eng. Syst.*, vol. 17, no. 3, pp. 417–427, 2024. doi:10.22266/ijies2024.0630.33.
- J. P. Thekkekara, S. Yongchareon, and V. Liesaputra, “An Attention-Based CNN-BiLSTM Model for Depression Detection on Social Media Text,” *Expert Syst. Appl.*, vol. 249, article 123834, 2024. doi:10.1016/j.eswa.2024.123834.
- M. Kerasiotis, L. Ilias, and D. Askounis, “Depression Detection in Social Media Posts Using Transformer-Based Models and Auxiliary Features,” *Soc. Netw. Anal. Min.*, vol. 14, article 196, 2024. doi:10.1007/s13278-024-01360-4.
- A. Qasim, G. Mehak, N. Hussain, A. Gelbukh, and G. Sidorov, “Detection of Depression Severity in Social Media Text Using Transformer-Based Models,” *Information*, vol. 16, no. 2, article 114, 2025. doi:10.3390/info16020114.
- N. Hussain, A. Qasim, G. Mehak, M. Zain, G. Sidorov, A. Gelbukh, and O. Kolesnikova, “Multi-Level Depression Severity Detection with Deep Transformers and Enhanced Machine Learning Techniques,” *AI*, vol. 6, no. 7, article 157, 2025. doi:10.3390/ai6070157.

A. Alambo, M. Gaur, U. Kursuncu, K. Thirunarayan, and A. Sheth, "Depression Detection from Social Media: Exploiting Affective and Contextual Information," in *Proc. IEEE Int. Conf. Healthcare Informatics (ICHI)*, 2019.

I. Baydili, B. Tasci, and G. Tasci, "Deep Learning-Based Detection of Depression and Suicidal Tendencies in Social Media Data with Feature Selection," *Behav. Sci.*, vol. 15, no. 3, article 352, 2025. doi:10.3390/bs15030352.

M. K. Effendi, Sriyanto, S. Y. Irianto, C. Fauzi, and Y. Vitriani, "Optimizing Student Depression Prediction Using Particle Swarm Optimization and Random Forest," *J. CoreIT*, vol. 11, no. 1, pp. 44–50, 2025. doi:10.24014/coreit.v11i1.35954.

Y. Wang, L. Xu, H. Qian, H. Lin, and X. Zhang, "Improved Empirical Wavelet Transform Combined with Particle Swarm Optimization-Support Vector Machine for EEG-Based Depression Recognition," *EURASIP J. Adv. Signal Process.*, vol. 2024, article 101, 2024. doi:10.1186/s13634-024-01199-z.

A. Atta, D. ElSayad, D. Ezzat, S. Amin, and M. ElGamal, "Efficacy of Swarm-Based Neural Networks in Automated Depression Detection," *Sci. Rep.*, vol. 15, article 25560, 2025. doi:10.1038/s41598-025-09414-z.

R. Geetha and D. Vimal Kumar, "Cockroach Optimized Progressive Laplace Extreme Learning Machine for Depression Prediction in Social Media Text," *SSRG Int. J. Eng. Trends Technol.*, vol. 74, no. 4, pp. 132–143, 2026. doi:10.14445/22315381/IJETT-V74I4P110.

S. Chitrakala and S. Senthamizhselvi, "GO-DEPRESSION: Gannet Optimization Codemix Features Based Coordinate Attention CNN-BiLSTM for Penta Depression Detection," *Int. J. Comput. Intell. Syst.*, vol. 19, article 111, 2026. doi:10.1007/s44196-025-01142-6.

S. A. Thamrin and A. L. P. Chen, "Enhancing Explainability and Performance of the Depression Detection Model on Social Media Utilizing Feature Engineering and LLMs," *Health Inf. Sci. Syst.*, vol. 14, article 48, 2026. doi:10.1007/s13755-026-00446-x.

H. Mao and Q. Han, "Enhancing TextGCN for Depression Detection on Social Media with Emotion Representation," *Front. Psychol.*, vol. 16, article 1612769, 2025. doi:10.3389/fpsyg.2025.1612769.

I. Arshad, S. Naseem, M. Mussarat, A. Zanib, S. Al-Otaibi, A. Yousif, T. Mahmood, and A. R. Khan, "A Multi-Platform Approach to Optimizing Depression Detection in Social Media Through Feature Engineering and Contextual Analysis," *Int. J. Comput. Intell. Syst.*, vol. 19, article 97, 2026. doi:10.1007/s44196-026-01225-y.

Kaggle, "Mental Health Social Media Dataset." [Online]. Available: <https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media>.

M. Gao, "Depression Tendency Processing Based on LSTM Technique for Text Emotion Recognition," *Trans. Comput. Sci. Intell. Syst. Res.*, vol. 7, pp. 216–223, 2024. doi:10.62051/23e9s737.

A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020.