# A STACKED ENSEMBLE TECHNIQUE WITH GLOVE EMBEDDING MODEL FOR DEPRESSION DETECTION FROM TWEETS

Reseena Mol N.A[1]

Research Scholar,Dept.Of Computer Science
Karpagam Academy of Higher Education
Coimbatore,India
reseena.n.a@gmail.com

Dr.S.Veni[2]

Professor,Dept.Of Computer Science.
Karpagam Academy of Higher Education
Coimbatore,India
venikarthik04@gmail.com

**Abstract**

**With the increasing volume of web content on social network sites like Facebook, Twitter, etc. identifying the attitude of people becomes an easy task. That attitude can be used as an input to find the mental status of that person through their texts. According to WHO, depression is a general mental disorder, which has already affected more than 264 million people. With the help of sentiment analysis, it is possible to detect depression at an early stage from their tweets as they represent their attitude. Machine Learning Classification algorithms help to classify the texts as Depressed or non-depressed, but their accuracy is limited when researchers are using only traditional Bag of Words vectorizers to extract features. Instead of this, word embedding models can be used which represent words as real-valued vectors in a distinct vector space that is already defined and provides better accuracy. In this paper, we try to implement eight machine learning techniques for depression detection from tweets namely SVM, Logistic Regression, ExtraTree, Bagging, Random Forest, AdaBoost, XG boosting, and Gradient Boosting which employs different word embeddings like Word2Vec, FastText, and Glove on the dataset which is used to train the classifiers and performed a comparative evaluation of word embedding models with different classifiers. This study proposes a new Stacked ensemble technique with a glove embedding model for Depression detection from Tweets which provides higher accuracy than standalone models.**

***Keywords:*** **Depression,Tweets,Word embeddings,Bag of Words,Vectorizer,Word2Vec,Glove,Fasttext**

## 1. Introduction

One of the universal mental disorders is Depression having a probable prevalence of 4.4% [2]. Depression affects the professional as well as the personal life of a person, reduces the number of working days, and increases the chances of illnesses like cardiac disease, diabetes, hypertension, etc. Depression is a curable mental illness, but a majority of the patients are not getting the chance of being treated on time because of many reasons. The traditional treatment requests direct interaction and the state of illness is not noticed at an early stage. But nowadays usage of social network sites is growing rapidly and people feel free to express their opinions through sites like Facebook, Twitter, blogs, etc. which offers people a chance to isolate their negative emotions at an early stage.

Sentiment analysis can be treated as a pooled procedure of text mining, NLP, and computational linguistics[3]. It helps to extract the sentiment concealed in the text-positive, negative or neutral. Twitter is a microblogging site where people are conveying their attitudes and opinions very freely about an event, product, etc. which helps the users to recognize their emotions. Tweets often represent a person's present state of mind, which can be positive or negative. The early detection of the undesirable mental state of a person indicates that they are inclined to have a mental disorder, sentiment analysis plays an important role in the detection of depression at the primary stage.

Machine learning approaches for sentiment analysis depend on machine learning algorithms by making use of syntactic as well as semantic features[6]. These techniques use different procedures for training the dataset to develop the classification model, the model created on this training dataset will be used for classification and also to check the correctness of classification with the help of the testing dataset. Machine learning approaches perform the operations in two ways- Supervised and Unsupervised. The supervised method is executed where the dataset has labeled data and the unsupervised method is used for unlabeled data.

For training the model, the text contained in the dataset has to be converted into vectors, an old-style bag of words vectorizers is simple but they are having large dimensions and are sparse. To avoid the inconveniences in handling these vectors and features word embeddings are introduced. Word embeddings are dense, distributed, fixed-length word vectors, built using word co-occurrence statistics as per the distributional hypothesis[7].

This paper experimentally compares three embedding methods Word2Vec, Glove, and FastText with eight machine learning approaches which include some ensemble methods, and proposed a new stacked ensemble technique with higher accuracy than individual models.

This paper is structured as follows. Related works in Section II; Methodology in Section III, Experimental Procedure in Section IV, and Experimental Results and Discussions described in Section V. In Section VI explained Proposed Methodology, and finally in VII we conclude the paper.

## 2. Related Works

[3]Here authors have proposed a method that comprises SVM and Adaboost DT classifiers to extract sentiments involved in Tweets. And from their experimental analysis, it is concluded that the combined method gives better accuracy.

[7]Authors have performed a survey on Word embeddings. Here the work describes recent techniques for maintaining fixed-length, dense, and distributed demonstrations of words. Word embeddings encode accurate syntactic and semantic word relationships. Explained in detail both the prediction-based and count-based word embedding models.

[15]In the proposed work, authors have stated that sentiment analysis can be either performed using the Lexicon approach, machine learning-based approach, or hybrid approach and identified the problem with the Lexicon approach ie. the accuracy of the Lexicon approach decrements with the increase in the size of the dictionary. Also explained the steps required for the sentiment analysis process using machine learning methods. NB, SVM, and DT are used here for explaining the process.

[18] Here machine learning model XGBoost is used to diagnose depression from the Dutch citizen dataset. In the dataset used, only 5% shows depressed cases while the remaining are Non-depressed cases. Since the dataset is imbalanced, the authors formed multiple samples by oversampling(O), under-sampling(U), over-under(OU) sampling, and ROSE sampling techniques. By implementing with XGBoost model, it is found that Over and OU sampling techniques provide better accuracy in the diagnosis.

[19]In this paper, the authors have done a comparative study between Machine learning models, Deep learning, and a hybrid of these two models. And they concluded that Deep learning models give better accuracy even though they are complex models, whereas machine learning models give a little bit lesser accuracy but they are easy to implement.

[20] Here the authors have used different Machine Learning models for detecting depression from social media content. Models used are LR, NB, RF, and SVM. From the results obtained, it is stated that LR shows higher performance.

[21]In this paper, a predictive model is designed to detect depression based on tweets by extracting the sentiment of Arabic text. Checked the performance of four classifiers-NB, AdaBoostM1, Random Forest & Liblinear, and Liblinear model produces higher accuracy of 87.5%.

[22]In this work several DNN architectures are applied for sentiment analysis and their performance is evaluated. Four publicly available datasets are utilized in this implementation. Different word embeddings like Word2Vec Glove are applied in neural networks like CNN, RNN, etc. Also implemented is an Ensembled network with a stack of several DNN models each trained with different word embedding methods. The final prediction is based on either the average of all the models or by majority voting. Concluded that researchers can select the most perfect DNN technique depending mostly on the application type.

[23]In the proposed work, the authors implemented word embedding models with classifiers to improve the efficiency of sentiment analysis. word2vec embedding model is used with a random forest classifier in this work and gets an accuracy of 81%.

[24]In this work sentiment analysis on the Amazon data set is performed to recognize the sentiment polarity of writing. Here a hybrid approach for classification is employed by combining SVM with Random forest and through experiment, it is found that the combined classifier provides better results than pure SVM or RF.

[26]In the work accuracy of the machine learning classifier employing, word embedding is compared with that of a bag of words vectorization. It is stated that word embedding helps to identify the semantics of the words even if they are not included in the training data set.

[27]Here authors extracted data from Twitter. Ensemble methods are used here for the classification process by combining word sense disambiguation and wordnet synsets to feature vector and concludes that ensemble methods outperform traditional classification approaches.

[28] In the proposed work, a classifier is developed by combining four classifiers, naive bayes, decision tree, multi-layer perceptron, and LR. This model is compared with individual classifiers and concludes that the new classifier provides better accuracy.
[29]The proposed work employed a word vector refinement model to Word2Vec and Glove models, by which the vector values will be closer for similar words and show a high difference in vector values for dissimilar words. Deep learning methods are used as classifiers and compared with the results by applying redefined word embedding models and concluded that Glove worked well with DAN, Bi-LSTM, and Tree-LSTM whereas CNN did well with CNN.
[30]In this paper, the authors have proved through their experiment that traditional vectorizers affect the accuracy of classifiers in sentiment analysis, so they used word embedding models,word2vec for generating high dimensional vectors. From the results, it is found that the accuracy of the classification process improves.

## 3. Methodology

### *3.1. Word embeddings*

It is the known fact that machine learning and deep learning algorithms are unable to treat strings or plain text as they exist, all words in the text have to be converted into numerical values to use in classification models, in conventional methods the words are represented as indices in a vocabulary[9]. One such representation namely One Hot Encoding, where each item is represented as a binary vector ie. all values are zero except the index of that item will be marked with one. Another Vector Space Model(VSM),the entire content is interpreted as a vector having the whole vocabulary as a dimension and weights of individual dimensions are the word frequencies within the document[8]. Its simplicity and robustness made them popular but its large dimensional size, similarity issues ie. they are not treating the words with similar features as matching and its computational issues that occur when dealing with large sparse values smashed the accuracy of classification models.
Word embeddings are useful illustrations of words and often lead to improve performance in the various tasks performed and used in almost all NLP(Natural Language Processing) tasks. They are maintaining valuable syntactic and semantic features of all the words. Some of the Predictor based on word embedding models are explained in this section.

#### *3.1.1 Word2Vec*

In one hot encode representation, all words are independent of each other. Objective of Word2Vec is to produce distributed representations, which exposes some dependence of one word on other words. Embeddings in Word2Vec is obtained with two models:- CBOW(Continuous Bag of Words) and Skip Gram. Both models are centered on the acquisition of words in their local usage scenario and the context is well-defined by a window of adjacent words.
Continuous bag of words models learns word embedding by estimating the current word following the context of the corpus. In contrast, the skip-gram model, acquire by predicting the context merely based on the current word.

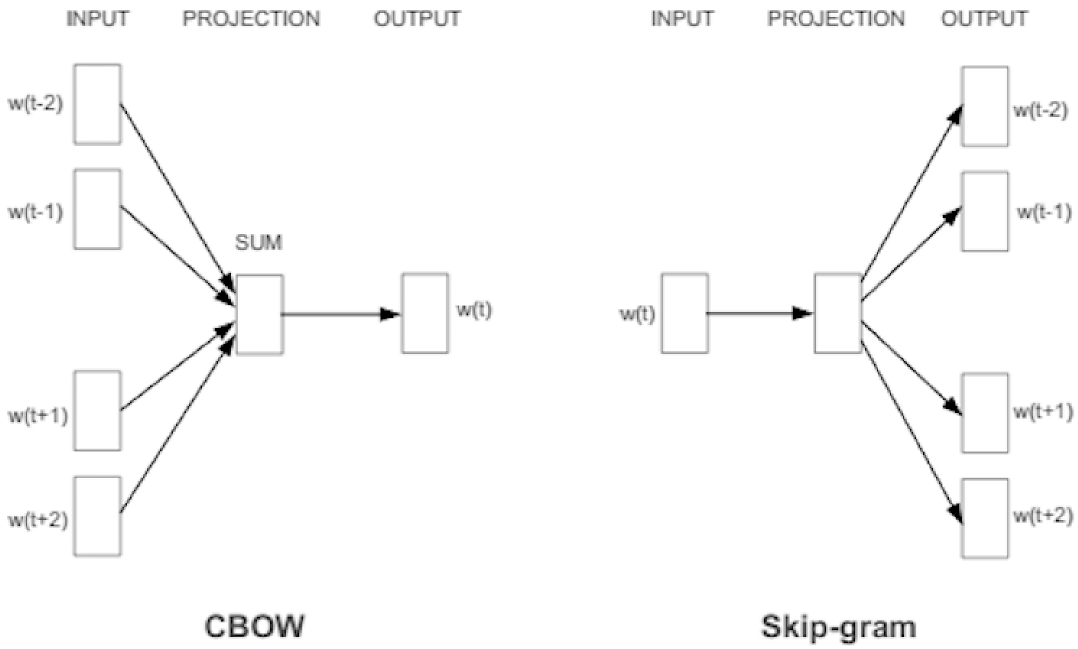


Fig. 1. Word2Vec Training Models[9]

### *3.1.2 Glove*

These are a type of word embedding that encrypts the overlapping probability ratio between two words as vector variances. The glove uses a weighted least squares objective J that reduces the variation between the dot product of the vectors of two words and the logarithm of their number of co-occurrences:

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

Equation 1. Weighting function of Glove model[10]

where wi and bi are the word vector and bias respectively of word i, $\tilde{w}_j$ and $\tilde{b}_j$ are the context word vector and bias respectively of word k, $X_{ij}$ is the number of times word i occurs in the context of word j, and f is a weighting function that assigns lower weights to unusual and continual co-occurrences.[10]

### *3.1.3 FastText*

The most prominent limitation of the Skip-gram embedding model is that they have separate vector representations for every individual word, not considering subwords in a single word. This is overwhelmed with the FastText embedding method. Its base is the Skip-gram model. In this model for each character n-grams, a vector depiction is associated with it, for example, the n-grams for the word “extract,” “ext,”, “rac” and ”act”, and then the word is represented as the sum of these vectors[11]. This model also represents words that are out of trained vocabulary[12]. This gives a chance of representing aptly the rare words as their n-grams may already be represented[13].

## ***3.2. Classifiers***

### *3.2.1 Machine Learning techniques*

In Machine learning techniques the sets of data are usually divided into two – training data set and testing data set. With training data set classifier generates a model ,which is then tested against the testing data set to ensure the essential accuracy. After attaining the required accuracy, this classifier can be used for predicting the label with real data. Machine learning algorithms used in this work are listed below.

3.2.1.1. Support Vector Machine(SVM)

The objective of SVM is to figure out linear separators or hyperplanes in search space to classify differently labeled data [15]. There can be several hyperplanes that do the same but the optimal separating hyperplane is one with the biggest margins ie. with space between the hyperplane and closest data point being maximum.

3.2.1.2. Logistic Regression(LR)

Logistic regression be associated with classifiers known as the exponential or log-linear classifiers[16]. This have similar nature with Naïve bayes as it extracts some weighted features from input, taking logs and taking the sum of product of weighted features.

3.2.1.3. Ensemble Techniques

They combines different base classifiers in order to generate a better one [17]. Usually they are used for improving the accuracy and efficiency of base machine learning classifiers. Generally NB, DT, SVM, and LR are used as base classifiers in ensemble classifiers. In this work,we have used Extra Tree, Bagging Tree, Random Forest, Adaboost, XG boosting and Gradient boosting ensemble techniques are used.

### *3.3. Performance measures*

After the classification model is developed, its accuracy is to be measured. Metrics considered for performance assessment are Accuracy, Precision, Recall and F1 score and these measures are based on true positivity and true negativity on result of classification process [17].

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$\text{Precision}=\frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{4}$$

$$\text{F1Score}=2\text{x}\frac{RecallxPrecision}{Recall+Precision} \tag{5}$$

## 4. Experimental Procedure

Comparative evaluation of different word embedding with classifiers for detecting depression from tweets involves the following tasks.

### *4.1. Dataset preparation*

Tweets are used here for the work, that has been collected on depressive and non depressive tweets from the publicly available Kaggle dataset. The dataset consist a total of 7013 tweets. Dataset preparation involves two tasks.

*4.1.1. Data Labelling*

To ensure the accuracy of the dataset, the help of a psychologist is sought, who takes each tweet individually and confirms the label as either depressive or non-depressive. Word cloud for depressive and non-depressive tweets for the data set is represented below.


Fig. 2 .Depressive word cloud


Fig. 3. Non-Depressive word cloud

### *4.1.2. Data preprocessing*

As the dataset consists of noises ,preprocessing is needed to guarantee precision. Preprocessing performed on the dataset was removing html links & mentions(#,@,https,: etc.),numbers, stop words, and punctuations. Performed Tokenization, lemmatizing and stemming operations to separate words and to get valid words from the tweets.

## *4.2 Feature Extraction*

### *4.1.3. Word embeddings*

For improved accuracy, Word Embedding models are used in this work, instead of just using vectorisers. Different embedding methods used are Word2Vec,Glove and FastText. BOW (Bag of Words) or TF-IDF will consider words separately as features ,but word embeddings represent real number vectors to represent semantic meaning of words and also similar word's representations will be close to each other.

With CBOW or Skip grams algorithm, Gensim library is used here to form word embeddings by training the own word2vec models on a custom corpus. Common parameters used to build vocabulary and to train the models are size=100(embedding's no. of dimensions), **window=5(**extreme space between a marked word and words around the same),**min_count=3(**smallest count of words to consider when training the model, words with an incidence less than this count will be ignored) and **workers=3**( The number of partitions through training and the default workers is 3)**.** For CBOW and Skip grams model, vectors are formed with two different methods- simple average and Tfidf weighting representations.

Glove pre trained model is downloaded and used as a model with 100 dimensions which is provided by Glove embedding model ,then converted into input to word2vec format using Glove word2vec function.

### *4.1.4. Vector Representaion*

Employed two methodologies to express tweet vectors. First method is the simple average (mean) of vectors of the words that constitute the tweet. In the equation, n for the number of words that make up the tweet, and it is different from one tweet to another:

$$\text{Vect}=\frac{\sum \text{Vectwordi}}{n} \qquad (6)$$

Equation 6. Vector values for tweet by average

And the second method is the mean Tfidf-weighting of the vectors, which is obtained by multiplying each vector by the Tfidf-weight of the corresponding word before calculating the average:

$$\text{TfidfVect}=\frac{\sum_{i=1}^{n} Tfidfweightword \; x \; Vectword}{n} \qquad (7)$$

Equation 7. Vector values for tweet by Tfidf weights

## *4.2. Classification process*

For the binary Classification problem, employed eight classifiers:SVM,Logistic Regression ,ExtraTree, Gradient boosting, Bagging, Adaboost, XGBoost and Random Forest. As previous works for detecting depression from tweets provides better results with these algorithms and with this work researchers are in search of finding the best combination of embedding model for these classifiers.

## 5. Experimental Results And Discussion

From the literature review done for this work, it is realized that most traditional Bag of words approaches for feature extraction was count vectorisers,one hot encode representation etc. but these methods are not suitable for many applications and they provide poor accuracy. For analysis purposes, count vectoriser for feature extraction is applied on Bagging classifier and got an accuracy of 81.38 for the result.
To find the best combination of the classifier with word embedding models, here we practically implemented eight different classifiers with four different embedding models- CBOW,Skip gram,Glove and Fasttext. For word2vec models ie. with CBOW and Skip gram models used the two vector representation methods, simple average mean method and Tfidf mean method. Evaluated accuracy of different classifiers and represented below the results in tabular form and also in the form of chart.

| **Classifier** | **Word Embedding models** | | | | | |
|---|---|---|---|---|---|---|
| | **CBOW Mean** | **CBOW Tfidf** | **SKIP GRAM Mean** | **Skip Gram Tfidf** | **Glove** | **Fasttext** |
| SVM | 75.12 | 82.75 | 92.02 | 91.09 | 92 | 69.85 |
| LR | 70.13 | 67.35 | 69.56 | 74.19 | 90.52 | 55.52 |
| Random Forest | 87.09 | 84.17 | 93.3 | 91.09 | 96.3 | 94.22 |
| Adaboost | 83.6 | 78.54 | 95.22 | 91.09 | 94.8 | 92.23 |
| Bagging | 87.31 | 84.81 | 93.58 | 90.87 | 95.1 | 93.3 |
| Gradient Boosting | 85.53 | 80.61 | 94.93 | 91.37 | 95.3 | 93.37 |
| XG Boost | 85.31 | 79.9 | 95.08 | 91.58 | 95.5 | 93.44 |
| Extra Tree | 86.52 | 84.03 | 93.65 | 91.3 | 96.6 | 94.79 |

Table 1.Accuracy measures of classifiers with different word embedding models

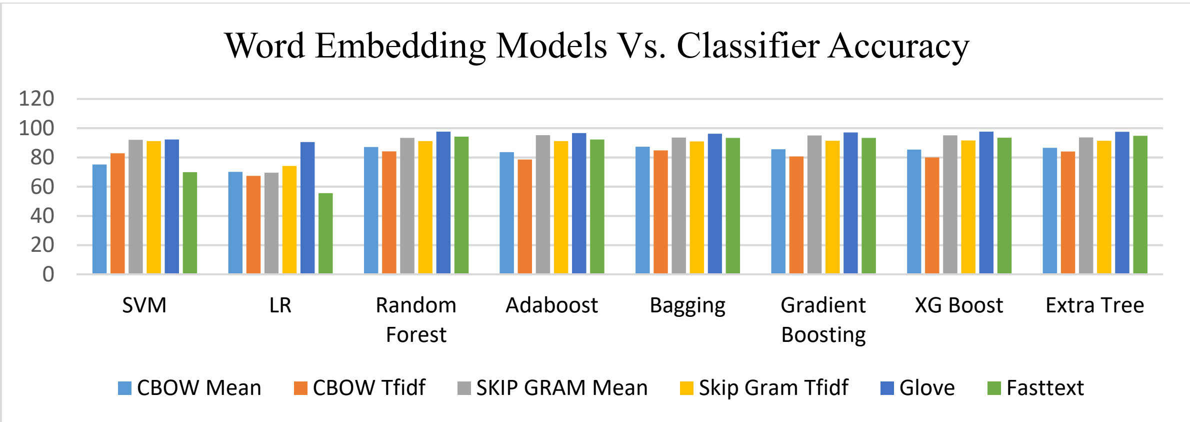


Fig. 4 Chart Representation Of Different Word Embedding Models With Classifier Accuracy Measure

Execution time is also analyzed through this experiment and represented the results below in the form of table and chart.

| Classifier | Word Embedding Models | | | | | |
|---|---|---|---|---|---|---|
| | CBOW Mean | CBOW Tfidf | SKIP GRAM Mean | Skip Gram Tfidf | Glove | Fasttext |
| SVM | 3 | 5 | 4 | 5 | 2 | 3 |
| LR | 1 | 3 | 1 | 2 | 1 | 1 |
| Random Forest | 7 | 11 | 8 | 11 | 9 | 8 |
| Adaboost | 12 | 16 | 13 | 15 | 14 | 13 |
| Bagging | 66 | 70 | 72 | 75 | 100 | 64 |
| Gradient Boosting | 31 | 34 | 33 | 34 | 33 | 31 |
| XG Boost | 5 | 9 | 6 | 8 | 8 | 6 |
| Extra Tree | 2 | 4 | 2 | 3 | 2 | 2 |

Table 2. Classifier's execution time with different word embedding models

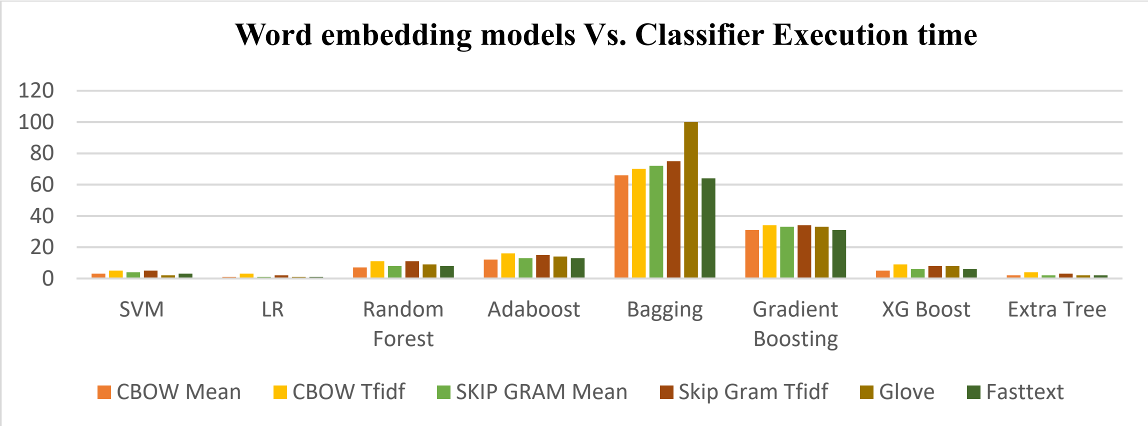


Fig. 5. Chart Representation Of various Word Embedding models with Classifier Execution time

In this work it is proven that traditional Bag of words feature extraction method, when count vectoriser is applied with Bagging tree classifier ,provides only 81.38 accuracy but when the same classifier uses word embedding methods for feature extraction offers higher accuracy.
Machine learning techniques SVM and LR gives lesser accuracy than other ensemble algorithms ,but they have the advantage of consuming lesser execution time.
From the results shown above, it is clear that all ensemble models do well in detecting depression and their accuracy barely varies for Glove & Fasttext word embedding models. It seems that Glove yields better results but the difference is very small. While considering execution time, all ensemble classifiers demand the practically same amount of time, but more time is needed when comparing it with SVM and LR.

## 6. Proposed Methodology – Stacked Ensemble Technique

From the experimental analysis, concluded that the Glove embedding model provides better accuracy with ensemble as well as with SVM and LR. So a stacked ensemble technique is developed for depression detection from tweets. Proposed method uses a stacked methodology to build an ensemble of Random Forest, ExtraTree, XGBoost and SVM classifiers. Stacking is an ensemble of diverse base learners and a meta learner which applies the output prediction of base learners as an input and then brings about final predictions [31][33][34]. The stacking model's structural design contains two or more fundamental models and a meta-model, which pools the estimates of base models. Fundamental models are known as Level -0 models and meta-model is branded as Level – 1 models. After training the base classifiers on complete training data, Meta learners are trained on the output of these base classifiers to best combine their results[32].

In the proposed method, used Random forest, Extratree, XGBoost, and SVM as base models while Logistic Regression is used as the meta-model. All the models involved in stacking use the Glove embedding models, as the experiment proves that Glove is the one that provides higher accuracy than all other embedding models. The Extra tree classifier gives better accuracy of 96.6 when all the classifiers are working independently. By stacking four of these models formed a single model and used a logistic regression model to associate the predictions from each of these four models in the most appropriate way. In this case, can see that the stacked ensemble method performs better than any single model on average achieving an accuracy of about 97 percent. A box plot is generated which shows the position of model classification accuracies.

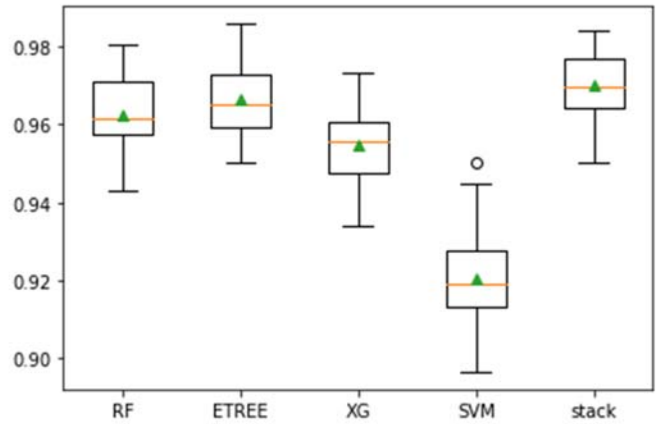


Fig. 6. Boxplot diagram for stand alone and Stacking Model Accuracies

## 7. CONCLUSION

In this work presented the approach of using different word embedding models for feature extraction with different machine learning classifiers for detecting depression from tweets. From the related works, it is found that the traditional Bag of words method for vectorizers is not doing great with classifiers, hence we implemented different word embedding models - CBOW, Skip Gram, Glove, and Fasttext with eight different classifiers to identify variance. From the results of the experiments, it is verified that the employment of word embedding models with classifiers causes more precise results for the problem and it is proven that employing the Glove embedding model with ensemble classifiers can be considered the better combination. From that conclusion developed a new Stacked Ensemble Technique for depression detection from tweets by combining four heterogeneous base models with Logistic regression as a meta-model and achieved a higher accuracy of 97 percent for the stacking model than the standalone models. We are planning to incorporate Deep learning techniques with a word embedding model to detect depression from tweets in future work.

## References

[1] P. Sharan, “An overview of Indian research in personality disorders,” Indian J. Psychiatry, vol. 52, no. 7, p. 250, 2010, doi: 10.4103/0019-5545.69241.
[2] B. B. Bifftu, W. W. Takele, Y. D. Guracho, and F. A. Yehualashet, “Depression and Its Help Seeking Behaviors: A Systematic Review and Meta-Analysis of Community Survey in Ethiopia,” Depress. Res. Treat., vol. 2018, pp. 43–47, 2018, doi: 10.1155/2018/1592596.
[3] A. Tariyal, S. Goyal, and N. Tantububay, “Sentiment Analysis of Tweets Using Various Machine Learning Techniques,” 2018 Int. Conf. Adv. Comput. Telecommun. ICACAT 2018, pp. 2–4, 2018, doi: 10.1109/ICACAT.2018.8933612.
[4] Maryam Mohammed Aldarwish ,Hafiz Farooq Ahmad,”Predicting Depression Levels Using Social Media Posts”, 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS), Volume: 1, Pages: 277-280.DOI Bookmark: 10.1109/ISADS.2017.41.
[5] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,” Knowl. Inf. Syst., vol. 60, no. 2, pp. 617–663, 2019, doi: 10.1007/s10115-018-1236-4.
[6] A. Gupta and J. Pruthi, “Survey on Sentiment Analysis for Twitter,” vol. 8, no. 03, pp. 51–60, 2017.
[7] F. Almeida and G. Xexéo, “Word embeddings: A survey,” arXiv, no. 1991, 2019.
[8] J. Camacho-Collados and M. T. Pilehvar, “From word to sense embeddings: A survey on vector representations of meaning,” arXiv, vol. 63, pp. 743–788, 2018.
[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” 1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc., pp. 1–12, 2013.
[10] HANSON ER, “Musicassette Interchangeability. the Facts Behind the Facts,” AES J. Audio Eng. Soc., vol. 19, no. 5, pp. 417–425, 1971.
[11] P. Association for Computational Linguistics, E. Grave, A. Joulin, and T. Mikolov, “Transactions of the Association for Computational Linguistics.,” Trans. Assoc. Comput. Linguist., vol. 5, pp. 135–146, 2017.

[12] C. Wang, P. Nulty, and D. Lillis, "A Comparative Study on Word Embeddings in Deep Learning for Text Classification," ACM Int. Conf. Proceeding Ser., pp. 37–46, 2020, doi: 10.1145/3443279.3443304.
[13] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, and J. F. Al Amri, "Evaluation of Sentiment Analysis via Word Embedding and RNN Variants for Amazon Online Reviews," Math. Probl. Eng., vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/5536560.
[14] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-Based Sentiment Analysis using {BERT}," Proc. 22nd Nord. Conf. Comput. Linguist., pp. 187–196, 2019.
[15] A. P. Jain and P. Dandannavar, "Application of Machine Learning Techniques to Mineral Recognition," Computer (Long. Beach. Calif)., no. October, pp. 628–632, 2016.
[16] M. Wankhade, A. Chandra, S. Rao, S. Dara, and B. Kaushik, "A Sentiment Analysis of Food Review using Logistic Regression," vol. 2, no. 7, pp. 251–260, 2017.
[17] Ankit and N. Saleena, "An Ensemble Classification System for Twitter Sentiment Analysis," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 937–946, 2018, doi: 10.1016/j.procs.2018.05.109.
[18] A. Sharma and W. J. M. I. Verbeke, "Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081)," Front. Big Data, vol. 3, no. April, pp. 1–11, 2020, doi: 10.3389/fdata.2020.00015.
[19] K. Jain and S. Kaushal, "A Comparative Study of Machine Learning and Deep Learning Techniques for Sentiment Analysis," 2018 7th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2018, pp. 483–487, 2018, doi: 10.1109/ICRITO.2018.8748793.
[20] G. Geetha, G. Saranya, K. Chakrapani, J. G. Ponsam, M. Safa, and S. Karpagaselvi, "Early Detection of Depression from Social Media Data Using Machine Learning Algorithms," ICPECTS 2020 - IEEE 2nd Int. Conf. Power, Energy, Control Transm. Syst. Proc., pp. 3–8, 2020, doi: 10.1109 / ICPECTS49113 .2020. 9336974 .
[21] S. Almouzini, M. Khemakhem, and A. Alageel, "Detecting Arabic Depressed Users from Twitter Data," Procedia Comput. Sci., vol. 163, pp. 257–265, 2019, doi: 10.1016/j.procs.2019.12.107.
[22] R. Wadawadagi and V. Pagi, Sentiment analysis with deep neural networks: comparative study and performance assessment, vol. 53, no. 8. Springer Netherlands, 2020.
[23] O. B. Deho, W. A. Agangiba, F. L. Aryeh, and J. A. Ansah, "Sentiment analysis with word embedding," IEEE Int. Conf. Adapt. Sci. Technol. ICAST, vol. 2018-Augus, pp. 1–4, 2018, doi: 10.1109/ICASTECH.2018.8506717.
[24] Y. Al Amrani, M. Lazaar, and K. E. El Kadirp, "Random forest and support vector machine based hybrid approach to sentiment analysis," Procedia Comput. Sci., vol. 127, pp. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.
[25] E. Rudkowsky, M. Haselmayer, M. Wastian, M. Jenny, Š. Emrich, and M. Sedlmair, "More than Bags of Words: Sentiment Analysis with Word Embeddings," Commun. Methods Meas., vol. 12, no. 2–3, pp. 140–157, 2018, doi: 10.1080/19312458.2018.1455817.
[26] M. Kanakaraj and R. M. R. Guddeti, "NLP based sentiment analysis on Twitter data using ensemble classifiers," 2015 3rd Int. Conf. Signal Process. Commun. Networking, ICSCN 2015, 2015, doi: 10.1109/ICSCN.2015.7219856.
[27] A. K. Abbas, A. K. Salih, H. A. Hussein, Q. M. Hussein, and S. A. Abdulwahhab, "Twitter Sentiment Analysis Using an Ensemble Majority Vote Classifier," J. Southwest Jiaotong Univ., vol. 55, no. 1, pp. 1–7, 2020, doi: 10.35741/issn.0258-2724.55.1.9.
[28] L. C. Yu, J. Wang, K. R. Lai, and X. Zhang, "Refining word embeddings for sentiment analysis," EMNLP 2017 - Conf. Empir. Methods Nat. Lang. Process. Proc., pp. 534–539, 2017, doi: 10.18653/v1/d17-1056.
[29] O. B. Deho, W. A. Agangiba, F. L. Aryeh, and J. A. Ansah, "Sentiment analysis with word embedding," IEEE Int. Conf. Adapt. Sci. Technol. ICAST, vol. 2018-Augus, no. March, pp. 1–4, 2018, doi: 10.1109/ICASTECH.2018.8506717.
[30] F. Divina, A. Gilson, F. Goméz-Vela, M. G. Torres, and J. F. Torres, "Stacking ensemble learning for short-term electricity consumption forecasting," Energies, vol. 11, no. 4, pp. 1–31, 2018, doi: 10.3390/en11040949.
[31] A. Wulamu ,B. Gaye, D. Zhang, and, "A tweet sentiment classification approach using a hybrid stacked ensemble technique," Inf., vol. 12, no. 9, 2021, doi: 10.3390/info12090374.
[32] Mustafa Temiz,Yasin Görmez, Yunus E. Işık, Zafer Aydın, "FBSEM: A Novel Feature-Based Stacked Ensemble Method for Sentiment Analysis' Comments in E-Government", International Journal of Information Technology and Computer Science(IJITCS), Vol.12, No.6, pp.11-22, 2020. DOI: 10.5815/ijitcs.2020.06.02
[33] N. S. Gill and S. Rani , "Hybrid model using stack-based ensemble classifier and dictionary classifier to improve classification accuracy of twitter sentiment analysis," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 7, pp. 2893–2900, 2020, doi: 10.30534/ijeter/2020/02872020.

## Authors Profile

**Reseena Mol N.A** received her Bachelor's in Computer Application from M.G.University,Kerala, India and her Master's in Computer Application from Indira Gandhi National Open University. She received her MPhil (Computer Science) from Vinayaka Mission University ,Salem. She is carrying out a research in Sentiment Analysis.

**Dr. S. Veni** received her Bachelor's in Computer Science, Master's in Computer Science,M.phil and P.hD from Bharathiar Uiversity. She is currently working as Professor and head, Department of CS, CA & IT, Karpagam Academy of higher education, India. She has successfully completed her Administrative responsibilities as Professor. Her research has included Control Systems, Data Analysis, Data Security.