Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 9 , 2025

AN ENHANCED ENSEMBLE LEARNING FRAMEWORK FOR STRESS DETECTION IN SOCIAL MEDIA TEXT USING LINGUISTIC ANALYSIS

Abdul Mueed [1] Talha Farooq Khan[2] Mubasher Malik[3]

[1, 2, 3] *Department of Computer science and Information technology, University of southern Punjab,Multan.*
[*]Corresponding Author: Abdul Mueed. Email: honeyraajput646@gmail.com

DOI: https://doi.org/10.5281/zenodo.17301004

Keywords
Ensemble Learning, social media, Stress detection, Linguistic analysis

Article History
Received 12 Aug 2025
Accepted on 29 Aug 2025
Published on 07 Sep 2025

Corresponding Author: *

Abdul Mueed

Abstract
This research presents a novel approach for the purpose of detecting stress using enhanced ensemble learning techniques combined with linguistic analysis, through social media text. With the growing issue of mental health concerns expressed online, accurate and timely detection of stress indicators is very important. The proposed model integrates multiple machine learning classifiers and improves detection accuracy by utilizing linguistic features such as syntax, semantics, and sentiment cues. A comprehensive dataset of SubReddit posts was used to train and evaluate the model. Results demonstrated that the ensemble approach outperformed individual classifiers and it is offering a promising tool for early stress detection and mental health monitoring in digital era.

INTRODUCTION

Mental Health is now a major issue of interest among individuals, communities as well as the government in all countries of the 21st century. Among other psychological problems, stress can be regarded as one of the most widespread and devastating issues that have prolonged effects on mental and physical condition. Refractory stress is linked to anxiety states, depression, sleeping disorders, heart diseases, and a poor quality of life. Conventional methods to stress detection have been more or less based on clinical interviewing, self-administered test questionnaires, and physiological surveillance but though these methods are effective, they are usually time-consuming, expensive, and inaccessible to big portions of the population, especially in low-income states or distant regions. As the internet and mobile technology expand at a very fast rate, social media has turned into a powerful medium of regular human communication. Real-time content that is shared on the social network by millions of users is based on their emotions, life experiences, and thoughts. This stream of data of such huge size, which is not organized in any way, is an excellent opportunity to acquire knowledge about the state of the psychology of users with the help of computational methods. More and more often, the theoretical possibilities to use Natural Language Processing (NLP) and machine learning to study the text on social media and identify stress and other mental disorders indicators are discussed by scholars. In contrast to the classic modalities, it is passive, scalable and even potentially real-time, and as such very suitable to the early warning model as well as population wide mental health surveillance.

Stress linguistically can be implicit, situational and culturally-responsive when written on social media. But it is usually expressed in certain language habits like the use of negative emotion words, words related to first-person pronouns, cognitive distortions, and syntactic variants. There are very expressive psychological reflections to examine these linguistic characteristics. Machine learning models particularly ensemble models which integrate the advantages of several classifiers have been promising in representing such patterns so as to effectively classify them. However, because of generalization difficulties, the problems of imbalance among classes, noisy data and the variation of human manifestation on various platforms, sometimes current models are challenging in order to address these challenges, the demand of stronger, flexible and intelligent systems is increasing. The proposed research suggests a modified version of ensemble learning incorporating using both traditional machine learning algorithms (SVM, Random Forest, and Naive Bayes), as well as deep learning such as BERT classifier. The framework will attempt to increase the accuracy and robustness of identifying stress in short, noisy, and informal text that appears in the social media by enriching the representation space with psychological, syntactic, and semantic features. This research would not only be added to computational linguistics and affective computing, but would also serve to contribute to the larger objective of creating automated tools to serve the purpose of mental health professionals, policymakers, and public health efforts.

In short, the study under consideration is one that falls under the study of psychology, linguistics and artificial intelligence and seeks to offer a solution to the ubiquitous problem of stress in the world by performing the new ensemble learning methods using the advent of twitters and other possible forms of social media. This issue can be reflected in the background of this work because it points to the significance of immediate, data-driven, and easily accessible mental health evaluation forms in the era of digitalization.

Hybrid Ensemble Framework: It introduces an innovative ensemble approach that integrates traditional algorithms (SVM, Naive Bayes, Random Forest) with deep learning models (LSTM) to enhance accuracy, stability, and generalization in stress detection.

Linguistic Feature Integration: It emphasizes the role of linguistic cues such as emotional tone, syntax, and stress-related keywords enabling interpretable and psychologically grounded predictions.

Practical Mental Health Application: The proposed model supports real-time stress detection from social media text, providing a foundation for online mental health monitoring and early intervention.

Interdisciplinary Impact: It bridges artificial intelligence and mental health research, offering methodologies applicable to related domains such as anxiety, depression, and emotional well-being.

- **Related work**

In the study(May & Ashik, 2024) the primary goal is stress detection by using Ensemble Learning Techniques such as SVM, DT, KNN, XGboost. The study revealed valuable insights into the field of stress detection. Ensemble accuracy achieved in the e study was 73%. The highest accuracy was achieved was from the Naïve Bayes. The use of sequential Neural Network enhanced the accuracy up to 76% and the total accuracy after voting and making the model ensembled the accuracy claimed in the research study is 73%. The dataset used in the research paper was downloaded from the Kaggle which is quite famous platform for machine learnings and deep learnings datasets. The dataset consists of 2838 of total instances and containing the subreddit posts labeled with stressed and not stressed where the stressed post's are labeled as 1 and non-stressed posts are labeled as 0.

In study (Febriansyah et al., 2022) the authors have developed stress detection system for the social media users. The techniques used are the traditional ML leaning like SVM, RF, Naive Bayes, a Decision Trees, Bag of Words Term frequency (TF) and inverse Term Frequency (IDF). The model achieved the accuracy of 75.00% and 80.00% and both were scored by SVM here were also some limitations too, such as smaller dataset could be used and usage of K-Fold on such models are suggested for better results.

The study (Philip et al., 2024) aimed to develop the Attention based model using CNN-BiLSTM for the detection of the level of depression. Deep learning models such as CNN-BiLSTM, BiLSTM attention, CNN BiGRU and the 0.89% F1 score was achieved along with accuracy of 96.71%. The author aims to include CLEF-2018 and CLEF-2019 datasets for future study.

Study (Tajuddin & Misbahuddin, 2020) demonstrated the Effective Stress Detection Framework (ESDF) by using Hybrid Ontology. The Technique quoted above is a keyword matching technique search process through which the stress of an individual is detected when he/she sends message to each other on social media. Probabilistic models such as GHS2 and the technique of Tree Alignment is used. The study lacks the usage of Hybrid Ontology / Multi lingual Language.

In the research (Illahi et al., 2022) paper the author has developed stress detection model by following the methodology such as data collection, data cleaning, splitting, boosting, bagging, voting and evaluation of the model. In this research ensemble machine learning approach is used such for stress detection in social media posts are classical techniques such as RF, DT, NB, SVM, Logistic Regression. The models achieved F1 Score of 76.60% with Logistic Regression. The limitations were seen such as lacking of BERT usage of single source of the data and smaller dataset.

The research article (Viktorovych et al.,2024.) is all about the practical implementation of the neural network's method for stress detection in social media posts which was carried out in the whole process. The methodology includes Pre-processing of the data its training / detection and formulation of conclusion. Accuracy of 0.95% was achieved along with Recall 0.94% with batch 64 and epoch 10.

The study (Tiwari & Das, 2021) shows the development of the model for stress detection based on machine learning classification only for tweets. The techniques were used such as KNN, SVM, D-trees, NB, TF-IDF, POS tagging as well. The study was based on the 1500 sentences extracted from the tweets. Ensemble approach gave F1 Score of 64 on all stress features.

The author (Guntuku et al., 2019) in concerned study used Happier Fun tokenizer with DLATK. It uncovers the insights for the language of the stress people. This study is majorly based on LWIC dictionaries which performed outstanding. The study lacks the casual insights as claimed by the author by stating that some people may stop using the social media while they are in stress and depression so in this case the stress goes unnoticed. This study also lacks the detection of stress specially dealing with the RAW posts. The future work and the suggestion include phone-based sensors.

The study (Vasha et al., 2023) is based on the traditional ML techniques such as SVM, RF, DT, LF, KNN and NB. The dataset was splitting is based on the criteria of 805 data for training the model and 205 for the testing of the model. BY using Catboost with LR, SVM, NB, RF, accuracy was 82.6% which was the best performing models in all the models used in the study.

The study (Wang et al., 2022) aimed to develop the no-lexicon based model framework in which the labeled dataset was used and accuracy was achieved about 0.68565 along with f1 Score of 0.6854 but in harder situation the model depicted different results such as the achieved accuracy was 0.9467 and F1 score was 0. 9465.The techniques include CAE, LTP, MTL, MAM.

The author (Smys & Raj, 2021) in the study developed the model for stress detection especially for tweeter users. The dataset which was used in the study is based on tweets which further consist of 2500 sentences. The Classification techniques used are quoted as SVM, NB, DT, RF. The words such as happy nice good were identified as positive emotions and words like hate nasty waste kill no-sleep ware concluded as negative or stress depicting words. The graph showed accuracy above 90%.

In a study (Kasmin et al., 2026) the goal of the author is to detect the stress through the text on social media platforms like Reddit and tweeter. It is well known by the people of the field of stress detection and NLP that these platforms are widely used in this purpose. The author has integrated the datasets of Reddit and Tweeter and used it for training the ML model. Different ratios of the data splitting is used in the first segment the data is split into the ratio of 70% for training 30% for the testing purpose then the ratio of 80% and 20% is used, lastly 90% data is used to train the data and only 10% data is used to test the model and its evaluation. It is clearly observed in the study that on each ration the modal gives slightly different results typically fall between 0.68 to 0.81 in which SVM attained the highest accuracy of 0.81%.

The goal of the study (Fatima et al., 2021) is prediction of stress by using NLP techniques and machine learning algorithms. ERT dataset is used which is based on Emotional Recall matched against Psychometric Scale such as DASS-21. Total 200 participants were included in the study/experiments and they were asked about the words for expressing their emotions.75% data was used to train the model and 25% data was kept to test and evaluate the model. The techniques which were used in the development of quoted model were DT, LSTM, RNN, MLP. The suicide notes were also included in the study for analysis which the main strength of this study as per my opinion. The results were based on Cross Validated prediction and the results were R=0.7 for depression, 0.44 for anxiety, 0.52 for stress. There were also some limitations in the study as it was claimed that LSTM and DT were unable to learn from the dataset used in the study as well as it cannot account for the content shifting. LASSO for semi-supervised regressions is claimed to be included in the study in future.

The study (Bharadwaj et al., 2025) depicts stress detection by using textual data from ML techniques such as LSTM, SVM, and Lexicon based approaches such as VADER which revealed promising prospectus for detection of stress using hybrid approaches i.e., Lexicon based approaches machine learning. There were some limitations found in the study such as Imbalanced dataset and multimodal data. It also lacks Real Time analysis for immediate support or remedy for stress which is crucial to avoid the large number of complexities at later stage.

2. Proposed methodology

In this section the complete methodology is outlined which is used for stress detection in social media text using hybrid approach such as using machine learning and deep learning models. This implemented framework is based upon pre-trained BERT, classical classifiers such as Naïve Bayes, Support Vector Machine and Random Forest. Weighted ensemble Strategy is used to combine/integrate the outcome which is yielded by all the models which are being used in the framework. In this chapter data pre-processing, feature engineering, model architecture, training strategies as well as ensemble methodology is elaborated.

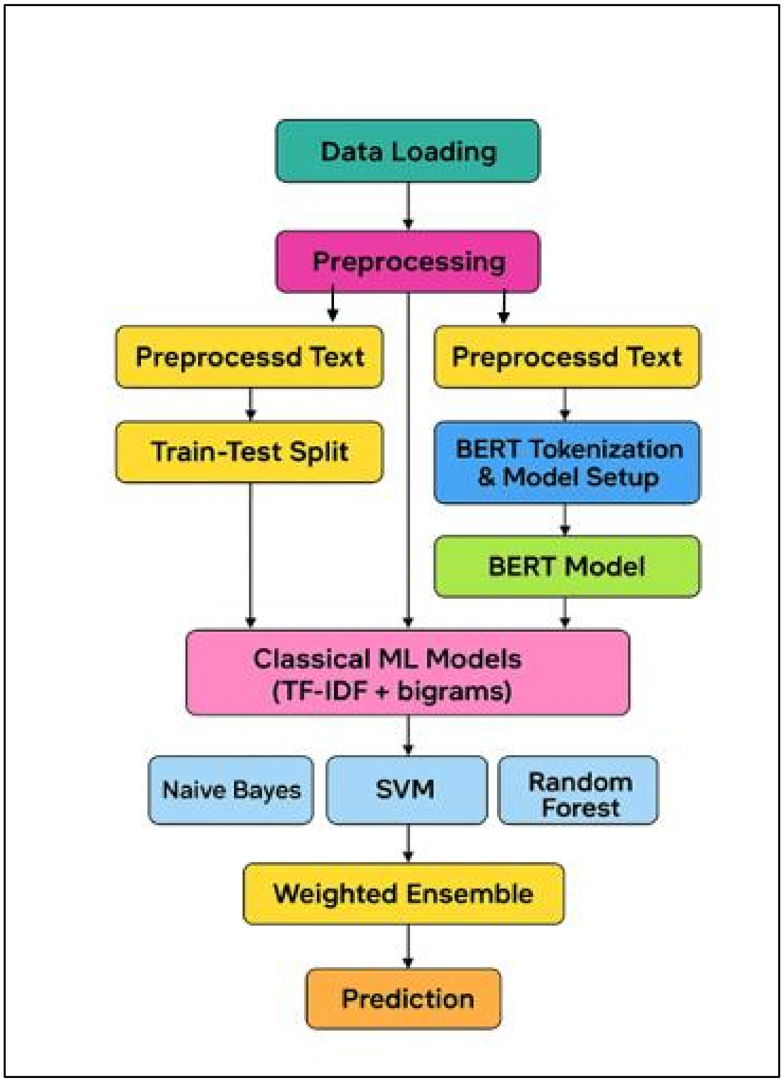

Figure 1 Methodology Diagram

3.1 Tools and Frameworks Used

The following tools and libraries were utilized in the framework:

- **Google Collab**: An online cloud based Jupiter notebook environment for developing, training and testing the models with support of GPU.
- **GPU**: T4 GPU with capacity of 15 GB virtual memory.
- **Python 3.8+**: Programming language purpose
- **Pandas, NumPy**: which is used for Data handling
- **TensorFlow & Keras**: Deep learning implementation
- **Transformers (HuggingFace)**: BERT pre-trained models
- **Scikit-learn**: which is used to train Classical ML models
- **Matplotlib**: used for Visualization

3.2 Experimental Setup

I. **Hardware:**

a) **System:** Dell Latitude 5480 (Internet-resource environment).

- **Ram:** 16 GB DD4.
- **Processor:** Core i5 8th Generation.
- **Rom:** 256 GB NVMe.
- **Batch Size:** 16 (optimized for limited memory).
- **Training Time:** Managed via early stopping and learning rate reduction.

3.3 Dataset Description

The dataset named as stress.csv is used in this research study, it consists of labeled data which includes social media text from Reddit which is downloaded from Kaggle, a well-known platform used for datasets and machine learning. The dataset is used to train the model for binary classification such as stressed and no-stress. The dataset is loaded by using CSV format. Following attributes are used to train the entire model.

Label classes: 0 indicates no stress and 1 indicates the textual data have stress features.

Text Field: A column which is named as text with the data of social media posts from Reddit.

Figure 2 Most Used Terms in the Dataset

We got a visual representation from Fig. 2 that clarifies the most used terms in the dataset we have used in our research study after integrating the figure that shows the word cloud of the column named as text. With the help of this visualization, readers can easily identify important patterns in the text and gain insights about the patterns and shape of the dataset.

Pre-Processing Techniques:

i. Removal of HTML tags, Special characters.
ii. Conversion of the text into lower case.
iii. Token filtering is used for removal of stop words.

3.4 Data Preprocessing

The dataset was uploaded by using Google Colab Environment and pre-processing of the data involved lowercasing, removal of HTML tags as well as punctuation removal and stop word removal. Two separate columns were prepared:

processed_text_ml: for traditional models
processed_text_bert: for BERT-based input

```
def preprocess_text_ml(text):
    if pd.isna(text): return ""
    text = text.lower()
    text = re.sub(r'<.*?>', '', text)
    text = re.sub(r'[^a-z\s]', '', text)
    words = [w for w in text.split() if w not in stopwords]
    return " ".join(words)

def preprocess_text_bert(text):
    if pd.isna(text): return ""
    text = text.lower()
    text = re.sub(r'<.*?>', '', text)
    return text
```

Figure 3 Pre-processing function

3.5 Feature Extraction and Classical Machine Learning Models

TF-IDF vectorization with bigrams was utilized/implemented for representation which Is textual in nature. The models which were trained are given below:

I. Naive Bayes (Multinomial NB)
II. Support Vector Machine (SVM)
III. Random Forest (RF)

Every model was trained using Scikit-learn pipelines to gain consistency.

3.5.1 Naïve Bayes

The Naïve Bayes is a probabilistic Classifier which is based on the Bayes Theorem. The quoted classifier does decision on the bases of probability of an event calculated through the Bayes Theorem. On the probability of an event which can be higher or lower, it will make decision and perform the task of classifying the text which can be stressed and not stressed.

```
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import make_pipeline
vectorizer = TfidfVectorizer(max_features=5000)
nb_model = make_pipeline(vectorizer, MultinomialNB())
nb_model.fit(X_train, y_train)
nb_probs = nb_model.predict_proba(X_test)[:, 1]
```

Figure 4 Code for Naive Bayes

ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 9 , 2025

The equation of Bayes theorem is given by:

$$P(A|B) = P(B|A) \cdot P(A) \quad /P(B)$$

Where:

- **P(A|B)**: The probability of event A occurring **given** event B has occurred (this is called the **posterior probability**).
- **P(B|A)**: The probability of event B occurring **given** event A has occurred (this is the **likelihood**).
- **P(A)**: The probability of event A occurring on its own (this is called the **prior probability**).
- **P(B)**: The probability of event B happening on its own (this is called the **evidence** or **normalizing constant**).

3.5.2 Supports Vector Machine

Support Vector Machine is a supervised learning algorithm that is effective in spaces which are high-dimensional. It works by finding the hyperplane that separates the data into different classes in a best way. SVM is always used for classification problems where the boundary for a decision is not linear as well as the data is sparse. In the provided code, a linear SVM is used to classify the data, with the following snapshot:

```
from sklearn.svm import SVC
svm_model = make_pipeline(vectorizer, SVC(kernel='linear', probability=True))
svm_model.fit(X_train, y_train)
```

Figure 5 Code snapshoot of SVM

3.5.3 Random Forest

The Random Forest is an ensemble learning method which is used for the purpose of classification when having multiple classes. It basically works by making multiple decision trees and choosing the majority prediction. Its enhanced accuracy factor is based upon majority decision due to which it becomes more reliable than as compared to a single tree. Random Forest is useful when the dataset is complex and single decision tree is not suitable and may not work well.

```
from sklearn.ensemble import RandomForestClassifier
rf_model = make_pipeline(vectorizer, RandomForestClassifier(n_estimators=100))
rf_model.fit(X_train, y_train)
rf_probs = rf_model.predict_proba(X_test)[:, 1]
```

Figure 6 Code for Random Forest

3.6 Text Embedding and Model Training

3.6.1 BERT Tokenization and Encoding

The BERT tokenizer was applied to prepare input ids as well as attention masks. These are used to feed in the model i.e. BERT.

```
def bert_encode(texts, tokenizer, max_len=128):
    input_ids, attention_masks = [], []
    for text in texts:
        encoded = tokenizer.encode_plus(
            text, add_special_tokens=True, max_length=max_len,
            padding='max_length', truncation=True,
            return_attention_mask=True, return_tensors='np'
        )
        input_ids.append(encoded['input_ids'])
        attention_masks.append(encoded['attention_mask'])
    return np.vstack(input_ids), np.vstack(attention_masks)
```

Figure 7 BERT Tokenization & Encoding

3.6.2 BERT Model Architecture

A custom BERT Layer was defined successfully by using TensorFlow's Keras API. CLS token embeddings were extracted and these were passed through dropout and dense layers and ultimately output with the help of a function named sigmoid which is used for decision making in binary classifications.

```
class BertLayer(Layer):
    def __init__(self, **kwargs):
        super(BertLayer, self).__init__(**kwargs)
        self.bert = TFBertModel.from_pretrained('bert-base-uncased')

    def call(self, inputs):
        input_ids, attention_mask = inputs
        outputs = self.bert(input_ids=input_ids, attention_mask=attention_mask)
        return outputs.last_hidden_state[:, 0, :]
```

Figure 8 BERT Layer

3.7 Weighted Ensemble

The Predictions from BERT and the three classical models were merged by using weighted ensemble approach. The Weights were based on all individual model accuracies on the validation set.

```
total_acc = nb_acc + svm_acc + rf_acc + bert_acc
w_nb = nb_acc / total_acc
w_svm = svm_acc / total_acc
w_rf = rf_acc / total_acc
w_bert = bert_acc / total_acc

ensemble_probs_weighted = (
    w_bert * bert_probs.flatten() +
    w_nb * nb_probs +
    w_svm * svm_probs +
    w_rf * rf_probs
)
ensemble_preds_weighted = (ensemble_probs_weighted > 0.5).astype("int32")
ensemble_acc_weighted = accuracy_score(y_test, ensemble_preds_weighted)
```

Figure 9 Weighted Ensemble approach

3. Result

It is a simple and fast model which is used with TF-IDF features.it works on the basis of Bayes theorem. In this study we used Multinomial Naïve Bayes which is proved effective when features are based on word counts. It is trained over the processed data using TF-IDF vectorizer. It is useful in identifying the important words while reducing the weight of common words. The model demonstrated accuracy of 0.738, precision of 0.893, and F1 score of 0.7734.

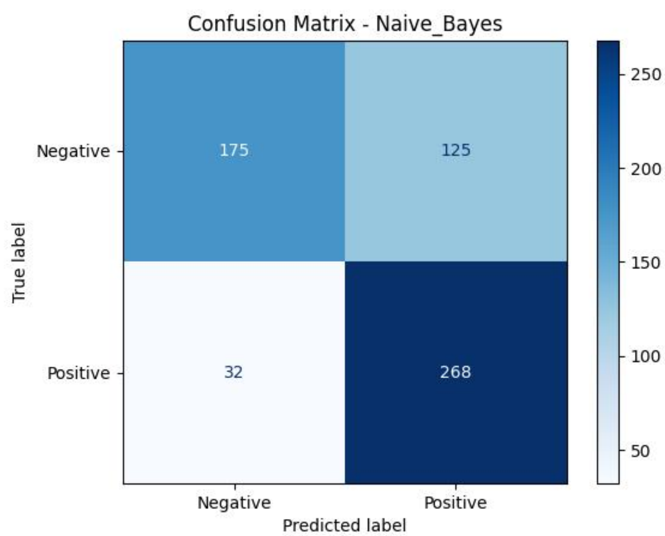

Figure 10 Confusion Metrics of Naive Bayes

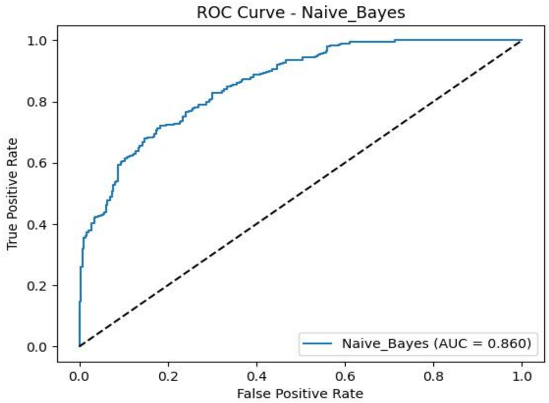

Figure 11 ROC Curve of Naive Bayes

4.2 SVM (Supports Vector Machine)

This model is very powerful and fast which is based on Supervised Machine Learning for tasks such as classification. It basically works by finding hyperplane which separates datapoints from different classes. In this study linear kernel SVM model with TF-IDF is used for the concerned task. SVM classifier demonstrates the accuracy of 0.7717, precision of 0.7540, recall of 0.7767 and F1 score of 0.7652.

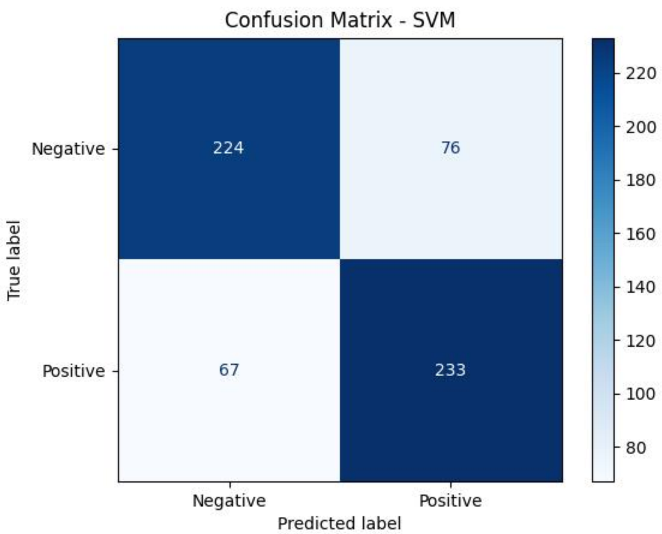

Figure 12 Confusion Metrics of SVM

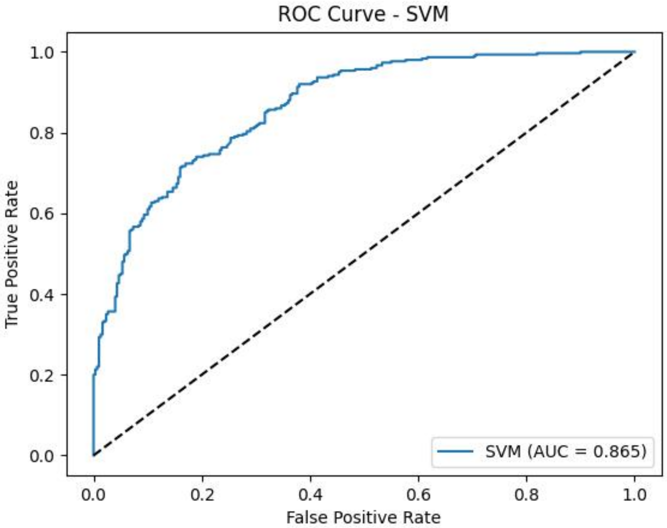

Figure 13 ROC Curve of SVM

4.3 Random Forest

It is a method which is based on ensemble learning for making decision trees specially when we deal with complex decision making. It works by concatenating their outputs. This performs best while handling the noisy data as well as the data which is high in volume Random Forest in this study yielded accuracy of 0.7417, precision of 0.7217, recall of 0.7867 and F1 score of 0.7528. This model shows that its performance is slightly lower as compared to SVM.

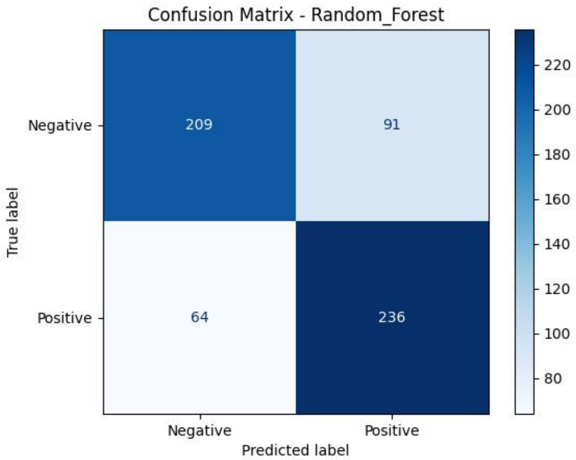

Figure 14 Confusion of Metrics of Random Forest

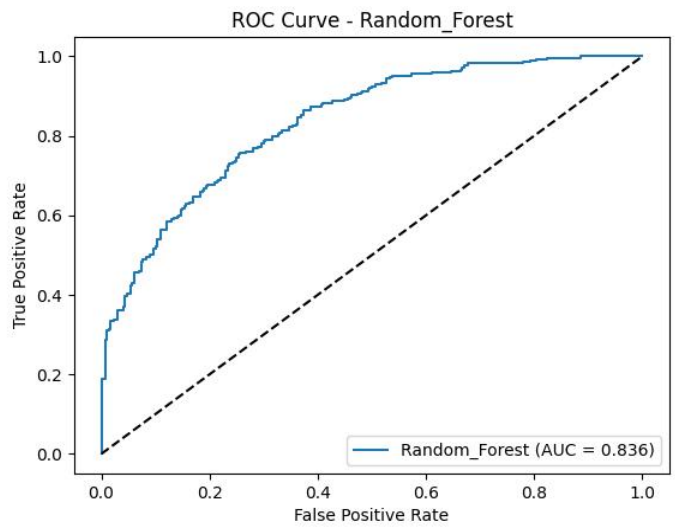

Figure 15 ROC Curve of Random Forest

4.4 BERT Classifier

BERT is the abbreviation of Bidirectional Encoder Representation from Transformers. It is based on deep learning as well as it is trained on large volume of textual data. It works by learning from the textual relationships between the words. This model was fine tuned for binary classification. The BERT model yielded the accuracy of 0.8117, precision of 0.8127, recall of 0.811 and F1 score 0.814. it was concluded that BERT outperformed all other traditional classifiers.

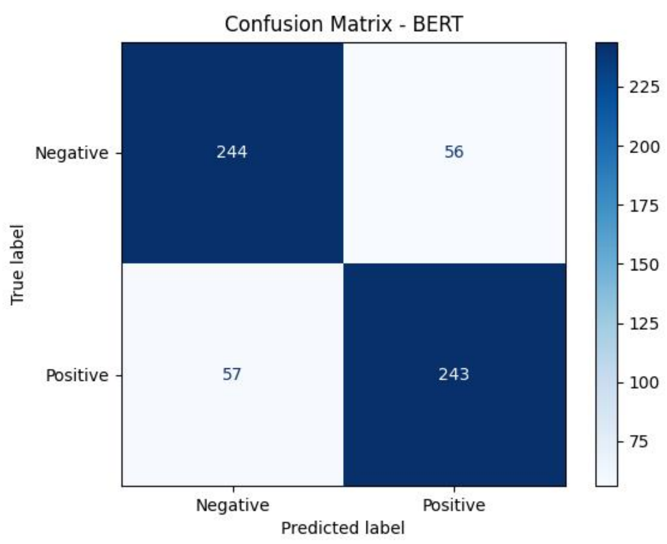

Figure 16 Confusion Metrics of BERT Classifier

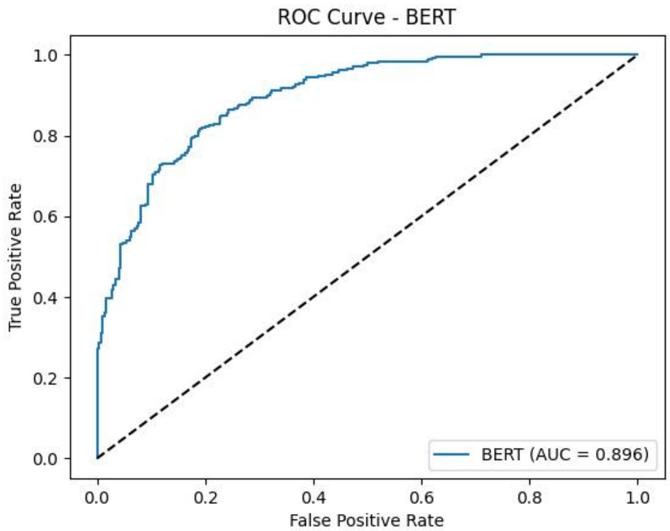

Figure 17 ROC Curve of BERT Classifier

4.5 Weighted Ensemble Approach

The weighted ensemble approach was used to combine the strength/outcomes of each models we used in this study. The weighted ensemble model achieved the performance by demonstrating the accuracy of 0.8133, precision of 0.7994, recall of 0.8367 and F1 score of 0.8176. The weighted ensemble model demonstrates that the accuracy or the performance of the overall model can be enhanced if we use weighted ensemble approach which combined the strengths/outcomes of all models.

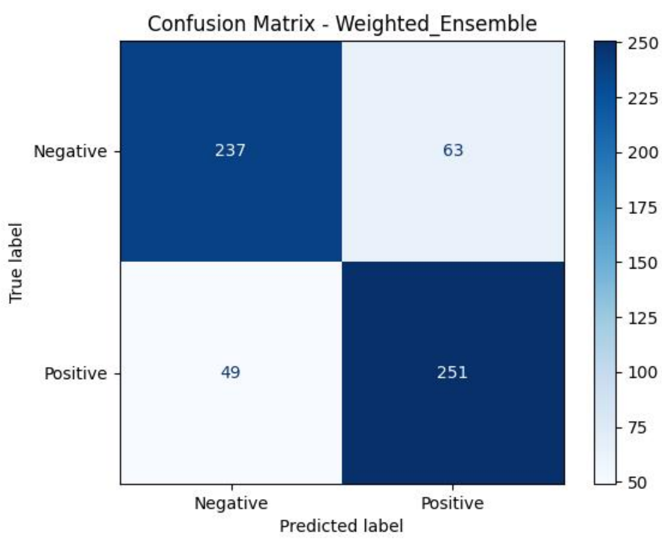

Figure 18 Confusion Metrics of Weighted Ensemble Model

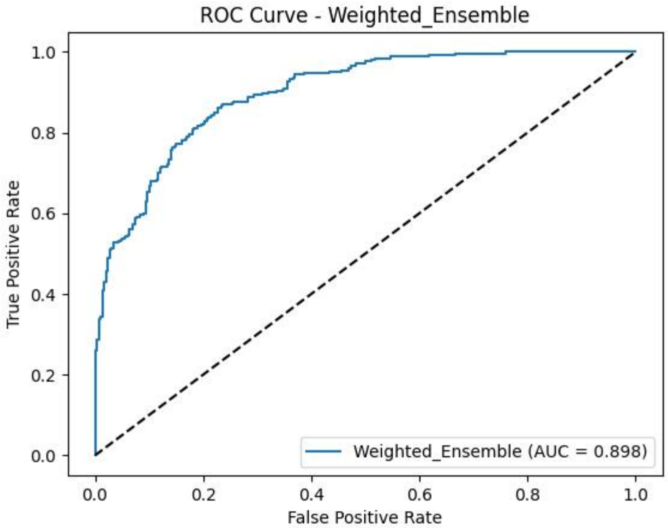

Figure 19 ROC Curve of Weighted Ensemble Model

Table 1 Comparison of all models in entire Framework

MODEL NAME	ACCURACY	PRECISION	F 1 SCORE	RECALL
SUPPORTS VECTOR MACHINE	0.7617	0.7540	0.7652	0.7767
NAÏVE BAYES	0.7383	0.6819	0.7734	0.8933
RANDOM FOREST	0.7417	0.7217	0.7528	0.7867
BERT	0.8117	0.8127	0.8114	0.8100
WEIGHTED ENSEMBLE	0.8133	0.7994	0.8176	0.8367

4. Conclusion

In this study different machine learning models are used to detect the stress by capturing the contextual relationships writing the textual data of social media posts to detect the stress of an individual for an earlier remedy to avoid further complications in the future. The study is based on the classical algorithms and deep learning approaches which were examined to perform text classification for detection of stressed posts. The models tested were Supports Vector Machine(SVM), Naïve Bayes(NB), Random Forest (RF), and BERT classifier a transformer-based model that revolutionized the NLP tasks in the field of NLP. Among all the classical models SVM showed balanced performance with accuracy, precision, recall F1 score. But it was seen that Naïve Bayes excelled in recall but did not perform comparatively well in precision. Random Forest gave satisfactory results but showed a little bit weak results in terms of precision as compared to SVM. In contrast, BERT outperformed the classical models. BERT achieved highest accuracy, precision, recall, and F1 score. Tribute to its ability to detect the complex relationship among the contextual textual data and relationships among the words. Furthermore, a weighted ensemble approach is used to combine the strengths of all the models being used in research study. This approach enhanced the results by weighting the contribution of each model's strengths compensating individual model's weaknesses, this approach resulted into highest overall accuracy, recall, precision, and F1 score. The ensemble method yielded superior results by combining outputs of all classical models as well as BERT a deep learning transformer-based classifier. The table which is given below depicts the comparison table of each model used to construct the entire framework for purpose of stress detection using linguistic analysis. Contrast to the classement of base paper (May & Ashik, 2024) using sequental neural network, which achieved an accuracy of 76% on the whole data, we was able to depict improved statistical results with this research several machine learning models as well as collection techniques. Our BERT model outperformed the base paper with an accuracy of 81.17%, and F1 score of 81.14%. Also, our Weighted Ensemble which is an algorithmic average of multiple models achieved an astounding accuracy level of 81.33%. This is much better than the ensemble model in base paper which was 73% Accuracy. The results highlight the effectiveness of our own model-based ensemble and transformer training methods in predicting stress leading this research to be a more sophisticated and reliable tool for predicting the stress with high accuracy. While the base paper popularized the potential of ensemble methods and sequential neural networks, our customized models BERT and the Weighted Ensemble are a robust alternative and have better generalization.

Reference

1) Eyob, L., Achamaleh, T., Tayyab, M., Sidorov, G., & Batyrshin, I. (2024). Stress recognition in code-mixed social media texts using machine learning. International Journal of Combinatorial Optimization Problems and Informatics, 15(1), 32.

2) Rahman, M. S., Ashik, A., & Rahman, M. S. (2024, May). Comprehensive Analysis of Stress Levels Using Ensemble Learning and Neural Networks for Insightful Understanding. In 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) (pp. 1344-1349). IEEE.

3) Agarwal, R., Rathore, R., Yadav, S., Jain, S., Singh, V., & Mishra, Y. (2019). Stress Detection of User using Social Interaction.

4) Ahmed, A., Aziz, S., Toro, C. T., Alzubaidi, M., Irshaidat, S., Serhan, H. A., ... & Househ, M. (2022). Machine learning models to detect

anxiety and depression through social media: A scoping review. Computer methods and programs in biomedicine update, 2, 100066.

5) AlArfaj, A. A., Hakami, N. A., & Hosni Mahmoud, H. A. (2023). Predicting Violence-Induced Stress in an Arabic Social Media Forum. Intelligent Automation & Soft Computing, 35(2).

6) Alghamdi, Z., Alatawi, F., Karami, M., Kumarage, T., Mosallanezhad, A., & Liu, H. (2023). Code RED: Reactive Emotion Difference for Stress Detection on Social Media. EasyChair.

7) Alghamdi, Z., Kumarage, T., Agrawal, G., Liu, H., & Bernard, R. (2024, May). Less is More: Stress Detection Through Condensed Social Media. In A Conference Hosted By.

8) Amanat, A., Rizwan, M., Javed, A. R., Abdelhaq, M., Alsaqour, R., Pandya, S., & Uddin, M. (2022). Deep learning for depression detection from textual data. Electronics, 11(5), 676.

9) Andrew, J. J. (2024, March). JudithJeyafreeda_StressIdent_LT-EDI@ EACL2024: GPT for stress identification. In Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion (pp. 173-176).

10) Ansari, L., Ji, S., Chen, Q., & Cambria, E. (2022). Ensemble hybrid learning methods for automated depression detection. IEEE transactions on computational social systems, 10(1), 211-219.

11) Athar, A., Mozumder, M. A. I., Fathima, K., Hussain, A., Ali, S., & Kim, H. C. (2023, September). Anxiety and depression detection and rehabilitation in the metaverse: a BERT-based recommendation system. In 2023 International Conference on Intelligent Metaverse Technologies & Applications (iMETA) (pp. 1-5). IEEE.

12) Ayyalasomayajula, M. M. T., Agarwal, A., & Khan, S. (2024). Reddit social media text analysis for depression prediction: using logistic regression with enhanced term frequency-inverse document frequency features. International Journal of Electrical and Computer Engineering (IJECE), 14(5), 5998-6005.

13) Bharadwaj, B., Nayak, S., & Panigrahi, P. K. (2025). Sentiment analysis for identifying depression through social media texts using machine learning technique. Big Data and Computing Visions, 5(2), 102-118.

14) Bobade, P., & Vani, M. (2020, July). Stress detection with machine learning and deep learning using multimodal physiological data. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 51-57). IEEE.

15) Cascalheira, C. J., Chapagain, S., Flinn, R. E., Klooster, D., Laprade, D., Zhao, Y., ... & Hamdi, S. M. (2024, May). The lgbtq+ minority stress on social media (missom) dataset: A labeled dataset for natural language processing and machine learning. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 18, pp. 1888-1899).

16) Chaware, S. M., Makashir, C., Athavale, C., Athavale, M., & Baraskar, T. (2020). Stress detection methodology based on social media network: A proposed design. International journal of innovative technology and exploring. Engineering, 9(3).

17) Meshram, S., Babu, R., & Adhikari, J. (2020, June). Detecting psychological stress using machine learning over social media interaction. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 646-649). IEEE.

18) Fatima, A., Li, Y., Hills, T. T., & Stella, M. (2021). Dasentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning. Big Data and Cognitive Computing, 5(4), 77.

19) Febriansyah, M. R., Yunanda, R., & Suhartono, D. (2023). Stress detection system for social media users. Procedia Computer Science, 216, 672-681.

20) Ghosh, S., Ghosh, S., & Das, D. (2017). Sentiment identification in code-mixed social media text. arXiv preprint arXiv:1707.01184.

21) Guntuku, S. C., Buffone, A., Jaidka, K., Eichstaedt, J. C., & Ungar, L. H. (2019, July). Understanding and measuring psychological stress using social media. In Proceedings of the international AAAI conference on web and social media (Vol. 13, pp. 214-225).

22) Guntuku, S. C., Buffone, A., Jaidka, K., Eichstaedt, J. C., & Ungar, L. H. (2019, July). Understanding and measuring psychological stress using social media. In Proceedings of the international AAAI conference on web and social media (Vol. 13, pp. 214-225).

23) Hossain, M. T., Talukder, M. A. R., & Jahan, N. (2021, July). Social networking sites data analysis using NLP and ML to predict depression. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1-5). IEEE.

24) Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

25) Zhuang, M., Cheng, D., Lu, X., & Tan, X. (2024). Postgraduate psychological stress detection from social media using BERT-Fused model. PloS one, 19(10), e0312264.

26) Illahi, M., Siddiqui, I. F., Ali, Q., & Alvi, F. A. (2022). Ensemble machine learning approach for stress detection in social media texts. Quaid-E-Awam Univ Res J Eng Sci Technol Nawabshah, 20(02), 123-128.

27) Inamdar, S., Chapekar, R., Gite, S., & Pradhan, B. (2023). Machine learning driven mental stress detection on reddit posts using natural language processing. Human-Centric Intelligent Systems, 3(2), 80-91.

28) Iswarya, P., Maheswari, D., & Rajeswari, R. (2020). Stress detection methods in Social networks: A Comparative study.

29) Jadhav, S., Machale, A., Mharnur, P., Munot, P., & Math, S. (2019, September). Text based stress detection techniques analysis using social media. In 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA) (pp. 1-5). IEEE.

30) Jussila, J., Alkhammash, E., Alghamdi, N. S., Madhala, P., & Khan, M. A. (2021). A netnographic-based semantic analysis of tweet contents for stress management. Computers, Materials and Continua, 70(1), 1845-1856.

31) Kasmin, F., Razali, N. A. I., Ahmad, S. S. S., Othman, Z., & Maylawati, D. S. A. Stress Detection Through Text in Social Media Using Machine Learning Techniques.

32) Kumar, A., Trueman, T. E., & Cambria, E. (2022, November). Stress identification in online social networks. In 2022 IEEE international conference on data mining workshops (ICDMW) (pp. 427-434). IEEE.

33) Kumari, K., & Das, S. (2022). Stress Detection System using Natural Language Processing and Machine Learning Techniques. In WNLPe-Health@ ICON (pp. 45-55).

34) Lin, H., Jia, J., Guo, Q., Xue, Y., Huang, J., Cai, L., & Feng, L. (2014, July). Psychological stress detection from cross-media microblog data using deep sparse neural network. In 2014 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.

35) Liu, D., Feng, X. L., Ahmed, F., Shahid, M., & Guo, J. (2022). Detecting and measuring depression on social media using a machine learning approach: systematic review. JMIR Mental Health, 9(3), e27244.

36) Rahman, M. S., Ashik, A., & Rahman, M. S. (2024, May). Comprehensive Analysis of Stress Levels Using Ensemble Learning and Neural Networks for Insightful Understanding. In 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT) (pp. 1344-1349). IEEE.

37) Nijhawan, T., Attigeri, G., & Ananthakrishna, T. (2022). Stress detection using natural language processing and machine learning over social interactions. Journal of Big Data, 9(1), 33.

38) Oryngozha, N., Shamoi, P., & Igali, A. (2024). Detection and analysis of stress-related posts in reddit's acamedic communities. IEEE access, 12, 14932-14948.

39) Thekkekara, J. P., Yongchareon, S., & Liesaputra, V. (2024). An attention-based CNN-BiLSTM model for depression detection on social media text. Expert systems with applications, 249, 123834.

40) Nijhawan, T., Attigeri, G., & Ananthakrishna, T. (2022). Stress detection using natural language processing and machine learning over social interactions. Journal of Big Data, 9(1), 33.

41) Jain, P., Srinivas, K. R., & Vichare, A. (2022). Depression and suicide analysis using machine learning and NLP. In Journal of Physics: Conference Series (Vol. 2161, No. 1, p. 012034). IOP Publishing.

42) Ramos, L., Shahiki-Tash, M., Ahani, Z., Eponon, A., Kolesnikova, O., & Calvo, H. (2024). Stress detection on code-mixed texts in dravidian languages using machine learning. arXiv preprint arXiv:2410.06428.

43) Rastogi, A., Liu, Q., & Cambria, E. (2022, July). Stress detection from social media articles: New dataset benchmark and analytical study. In 2022 international joint conference on neural networks (IJCNN) (pp. 1-8). IEEE.

44) Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. Journal of King Saud University-Computer and Information Sciences, 34(10), 9564-9575.

45) Saleem, M., & Kim, J. (2024). Intent aware data augmentation by leveraging generative AI for stress detection in social media texts. PeerJ Computer Science, 10, e2156.

46) Shelke, N., Chaudhury, S., Chakrabarti, S., Bangare, S. L., Yogapriya, G., & Pandey, P. (2022). An efficient way of text-based emotion analysis from social media using LRA-DNN. Neuroscience Informatics, 2(3), 100048.

47) Smys, S., & Raj, J. S. (2021). Analysis of deep learning techniques for early detection of depression on social media network-a comparative study. Journal of trends in Computer Science and Smart technology (TCSST), 3(01), 24-39.

48) Smys, S., & Raj, J. S. (2021). Analysis of deep learning techniques for early detection of depression on social media network-a comparative study. Journal of trends in Computer Science and Smart technology (TCSST), 3(01), 24-39.

49) Tajuddin, M., Kabeer, M., & Misbahuddin, M. (2020, January). Analysis of social media for psychological stress detection using ontologies. In 2020 Fourth International Conference on Inventive Systems and Control (ICISC) (pp. 181-185). IEEE.

50) Taspinar, Y. S., & Cinar, I. (2024, April). Stress detection with natural language processing techniques from social media articles. In Proceedings of international conference on intelligent systems and new applications (Vol. 2, pp. 70-74).

51) Turcan, E., & McKeown, K. (2019). Dreaddit: A reddit dataset for stress analysis in social media. arXiv preprint arXiv:1911.00133.

52) Turukmane, A. V., Avula, L. N., Gunji, C., Devapatla, R. N. K., & Ambati, S. (2025, May). Stress Detection Using NLP and DL Models. In International Conference on Sustainability Innovation in Computing and Engineering (ICSICE 2024) (pp. 1776-1788). Atlantis Press.

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

54) Mazurets, O. V., Sobko, O. V., Molchanova, M. O., Zalutska, O. O., & Yurchak, A. V. (2024). Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts.

55) Wan, X., & Tian, L. (2024). User stress detection using social media text: A novel machine learning approach. International Journal of Computers Communications & Control, 19(5).

56) Wang, X., Cao, L., Zhang, H., Feng, L., Ding, Y., & Li, N. (2022, April). A meta-learning based stress category detection framework on social media. In Proceedings of the ACM Web Conference 2022 (pp. 2925-2935).

57) Wang, X., Zhang, H., Cao, L., Zeng, K., Li, Q., Li, N., & Feng, L. (2023, August). Contrastive learning of stress-specific word embedding for social media based stress detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 5137-5149).

58) Yang, Y. C., Xie, A., Kim, S., Hair, J., Al-Garadi, M., & Sarker, A. (2023). Automatic detection of twitter users who express chronic stress experiences via supervised machine learning and natural language processing. CIN: Computers, Informatics, Nursing, 41(9), 717-724.