Stress Detection from Social Media Posts Using Sentiment and Emotion Analysis

Abdul Hye
Computer Science & Engineering
BRAC Univeristy
Dhaka, Bangladesh
abdul.hye@g.bracu.ac.bd

Rifah Jahan Tamanna
Computer Science & Engineering
BRAC Univeristy
Dhaka, Bangladesh
rifah.jahan.tamanna@g.bracu.ac.bd

Adiba Rahman
Computer Science & Engineering
BRAC Univeristy
Dhaka, Bangladesh
adiba.rahman1@g.bracu.ac.bd

Faysal Mohammed Abir
Computer Science & Engineering
BRAC Univeristy
Dhaka, Bangladesh
faysal.mohammed.abir@g.bracu.ac.bd

***Abstract*—Academic, professional and social pressures on society have made stress a significant mental health issue in modern society. Text-based automatic detection of the stress can facilitate early intervention and tracking of mental health. In this paper, a machine learning-powered technique used to identify stress in textual data has been presented to apply text processing methods of Natural Language Processing (NLP). Term Frequency – Inverse Document Frequency (TF -IDF) was used to preprocess and transform text data. The Logistic Regression and the Random Forest classifiers were trained and assessed based on the accuracy, precision, recall, F1-score, and ROC-AUC standards. Experimentally, the random Forest classifier has been found to be much better than the Logistic Regression in classification of stress. The suggested system proves that the classical machine learning models could be useful to detect the language patterns related to stress.**

***Index Terms*—Stress Detection, Natural Language Processing, Machine Learning, TF–IDF, Random Forest.**

I. Introduction

The rate of mental health disorders, especially stress, is on the rise all around the world and is extremely efficacious to productivity and wellness. Stressful conditions are usually shown in forms of language, and therefore textual analysis can be an effective tool in identifying it. As the online communication platforms continue to grow, there is a lot of textual information that can be analyzed automatically.

This paper will present a machine learning pipeline to identify textual stress using the NLP methods. It takes raw text and then preprocesses it, extracts features and then processes it using supervised classification algorithms. The goal of the project will be to consider the abilities of various machine learning models to detect stress.

II. Background and Literature Review

Recent developments in machine learning (ML) and natural language processing (NLP) have brought a great change in the way stress, anxiety, and depression are detected as the focus shifted to not only the conventional physiological indicators but also the data-driven and linguistic based and thinking modeling. The recent studies on this field as a whole are approached along three key directions namely multimodal physiological analysis, emotion-conscious social media modeling, and cognitive-semantic textual modeling.

Nayak et al. [1] provided an extensive overview of ML-based stress recognition based on physiological, behavioral, speech and text data. They had shown that the accuracies of Random Forest classifiers were able to reach up to 89 per cent on various datasets like DEAP and WESAD and were able to beat Support Vector Machines and Logistic Regression. Deep learning networks, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) performed more than 90 percentage when physiological measurements were used. The authors pointed out the issues of class imbalance, interpretability, and data privacy, and multimodal fusion and explainable AI were proposed as future directions of the field.

Alghamdi et al. [2] investigated affective and toxicity-based characteristics, as stress detection, on social media through a transformer-based pipeline composed of RoBERTa and Detoxify. Their Gradient Boosting model realised a 2.14percentage and 1.85percentage improvement in accuracy and F1-score over sentiment-only baselines respectively. The correlation analysis indicated that there was a strong positive association between stress and negative emotions like sadness and fear, but negative sentiment associated with positive.

Fatima et al. [3] proposed a model called DASentimental that is a model that fuses cognitive network science with machine learning together to identify that the text is in a depressed, anxious, or stressed shape. Their Multilayer Perceptron with an emotional recall data (in accordance with the DASS-21 scores) showed that this tool is strongly psycholinguistically valid as the correlation coefficients of depression, anxiety, and stress were 0.70, 0.44, and 0.52.

The scoping review presented by Ahmed et al. [4] uncovered

54 articles conducted on the subject of anxiety and depression detection based on social media and states Twitter and Reddit as key sources of data. They came to the conclusion that such ML models as SVM, Random Forest, LSTM, and CNN are practical but had certain ethical considerations concerning privacy and the transparency of their algorithms.

Wan and Tian [5] introduced a BiLSTM model with focus on detecting stress with the help of Dreaddit data and obtained an F1-score of 81.21 percentage. Their study proved the efficacy of attention mechanisms in attaching the language patterns which were stress-indicative.

Nijhawan et al. [6] fused sentiment analysis, emotion classification and topic modeling to predict the occurrence of stressors based on tweets at a 97.78 percentage accuracy with the help of the Random Forest and a fine-tuned BERT model of 94 percentage accuracy. They also highlighted that sentiment polarity is not enough to create accurate stress detection.

Rastogi et al. [7] provided new benchmarks and datasets on stress detection and demonstrated that transformer-based models are better in terms of performance compared to lexical and embedding-based models. Nonsense, however, automated annotation added the possibility of labeling noise.

Zhuang et al. [8] suggested a BERT-combined framework to combine topic modeling and BiLSTM-CRF to detect the postgraduate stress which attained F1-scores of over 92 percentage. The limitation of the study was that of one demographic and language, although it was effective.

According to this set of studies, there is a shift in the traditional ML models towards transformer-based and cognitively informed studies. Regardless of the tremendous advances in accuracy, there are still issues when it comes to the diversity of data, its interpretability, privacy and real time implementation.

III. Stated Problem and Objectives.

A. Problem Statement

The main research question that shall be examined in our research is the automatic classification of a textual case T, as found in the social media (posts on Reddit), into either one of two mutually exclusive categories: Stressed, which is marked with a positive class label 1; or Not Stressed, which is marked with a negative class label 0. It is a task that should be tackled through machine-learning algorithms that incorporate heterogeneous feature representations, thereby maximizing predictive accuracy while maintaining model interpretability, which is a requirement for substantive validation of mental-health applications.

The problem formulation is marred by an array of daunting issues. [3] First, the linguistic variation is the consequence of a variety of manifestations of stress among people, in different situations, and in different cultural milieus, which hampers finding universal textual signatures. Second, stress indicators are often subtle and implicit and require the use of models that are sensitive to the subtle use of language. Third, contextual dependency means that the same lexical items can have a stressful connotation in a certain situation but be neutral in other cases, which requires advanced contextual modelling. Fourth, the real-world corpora can also be imbalanced in the classes of the stressed and non-stressed samples, and this can skew the outputs of the classifiers. Fifth, the abundance of sarcasm, irony, and other figurative means only complicates the direct writing interpretation.

B. Objectives

The current study has a sequence of objectives, which it will seek to achieve using a systematic approach. The first goal is to maintain a high-quality, labelled stress-detection dataset, the Dreaddit corpus, and to execute a refined text-preprocessing pipeline, removing noise, URL, and other special characters, normalising, and linguistic format normalisation, in order to be able to ensure a consistent downstream processing.

The next target is the multimodal feature extraction. It consists of deriving TF -IDF vectors to extract salient lexical features, sentiment valence with the VADER lexicon - optimised to social-media discourse, and fine-grained emotional features with the DistilRoBERTa transformer. Also, Sentence-BERT is used to create semantic embeddings that are helpful in understanding the context.

The third objective consists of model development and optimisation. A logistic regression model is used as a baseline that can be interpreted, and random forest ensembles are built to improve the performance with the help of the combination of several decision trees. Both the modelling approaches have their hyper-parameter optimisation to achieve optimal predictive performance.

The fourth goal is that of thorough assessment. To be able to determine the accuracy, precision, recall, F1-score, and AUC metrics, the model is evaluated based on a set of metrics, combining accuracy, precision, recall, F1-score, and threshold-independent discrimination, respectively. The cross-validation processes are used to ensure robustness and careful error analysis is done to determine certain failure modes.

The fifth objective is the interpretability analysis that aims to explain the most significant attributes which are the basis of stress detection. This is obtained by performing feature-importance ranking, creating visualisations of word-clouds to represent lexical patterns between stressed and non-stressed text and analysing sentiment and emotion distributions to show affective patterns of stressed conversation.

The last goal summarizes the deployment of a pipeline. This pipeline offers an interface over which one can predict in real time with a new textual input and continuously stores learned models in production-ready formats, and provides visualisation tools, which make it easy to intuitively interpret model outputs.

IV. METHODOLOGY

The general outline of the system design is data preprocessing, feature extraction, model training, and evaluation.

A. Dataset Description

We use the Dreaddit corpus, a filtered corpus specifically designed to study psychophysiological studies related to the detection of stress in online social media speech. The corpus, which was conceived by Turcan and McKeown (2019), collects the submissions of the Reddit subgroups related to stress in five stress-relevant subgroups, including r/anxiety, r/stress, r/PTSD, r/almosthomeless, and r/assistance. This aggregate collection has a total of around 2,838 postings which are marked with binary values of absence (0) or presence (1) of stress. There is relative parity in the distribution of the labels, with stressed cases making about 55% and non-stressed cases making about 45%, and that is why the extensive interventions are not necessary to overcome class imbalance.

The length of texts differs significantly, as some may consist of short and one-sentence phrases, whereas others may be multi-paragraph accounts of aversive experiences. Everything is written in English, and commenting was done by human raters who evaluated the existence of stressful signs in the textual content.

B. Data Preprocessing

An elaborate preprocessing program is essential both to achieve a stable model behavior and to obtain features of a clean, normalised text representation. We have a pipeline that performs a series of operations to transform new submissions made to Reddit into an analyzable format.

To start with, all the text is changed into lower case to provide consistency and to avoid the model considering homographs with different capitalizations as different lexical items. This is followed by removing URLs by regular expressions, which recognize the prefixes of http and www and use auxiliary patterns to remove email addresses and any remaining HTML code that might be there. The third step eliminates unnecessary characters but still maintains simple punctuations that might carry an affective or emphatic message. Fourth, whitespace is normalised by replacing successive whitespaces with a single whitespace, as well as removing whitespace at the beginning and the end. Last but not least, we filter out the empty strings that might have appeared in the course of preprocessing and filter out the posts with fewer than three tokens, since the content of such short postings is often insufficient to make the classification credible.

C. Feature Extraction

Four different modalities of feature extraction are embraced in our multi-modal architecture that is intended to represent different linguistic and semantic aspects of stress expression.

First, TF -IDF elements generate 5,000-dimensional vectors. The TF -IDF - an acronym of Term Frequency-Inverse Document Frequency - adds more weight to the terms that are informative in the corpus by under-weighting common vocabulary. Its vectorizer is set with up to 5,000 features to trade off between exhaustive representation and computational tractability, document frequency minimum of 2 to reject incidental words, document frequency maximum of 0.8 to avoid over-represented words in single documents, n-gram range of one to two to represent both unigrams and bigrams and sublinear scaling of the term frequency to reduce the effect of over-represented words in individual documents.

Second, Sentiment analysis dimensions provide four dimensions through the VADER methodology (Valence Aware Dictionary and Sentiment Reasoner). VADER is a social media text customized, and it is capable of handling emoticons, colloquialisms, and intensifiers, which are characteristic of informal online conversations. The result is a negative score, a neutral score, a positive score, and a compound score which combines the overall sentiment on a normalized scale between -1 (most negative) and +1 (most positive). Compared to non-stressed samples, stressed samples are expected to record high negative scores and low compound scores.

Third, Emotion recognition features provide seven dimensions via DistilRoBERTa, which is a distilled version of the RoBERTa transformer that has been applied to the classification of emotions. This model gives out probability scores of the anger, disgust, fear, joy, neutral, sadness, and surprise emotions. Fine-grained emotion detection provides substantive granularity to binary sentiment, allowing the difference between adverse emotional valences, like fear and anger or sadness and disgust emotions, with potentially different relations to stress. We assume that the probability of fear, sadness, and anger will increase and the probability of joy will decrease in stressed texts.

Fourth, the semantic embeddings that BERT provides are 384 dimensions that are produced by SentenceBERT, namely the all-MiniLM-L6-v2 variant. Such dense embeddings capture semantic meaning and contextual relationships that can be missed by a sparse bag-of-words feature. The embeddings are provided with a strong ability to identify similarity of semantics and process synonyms through pre-training on large datasets that include both paraphrase detection and natural language inference, providing a complement to explicit lexical information obtained using TF-IDF.

The four sets of features are concatenated to give a single feature set to each sample with an overall dimensionality of 5,395 (5, 000 4 7 384). This is achieved by integrating sparse representations and dense representations, which are later encoded in sparse format in order to optimize the use of memory and computational abilities before final horizontal stacking.

D. Model Implementation

Two learning algorithms with supervision were put into practice to compare the relative effectiveness of the two in the detection of stress. The initial algorithm is the Logistic Regression, which is a linear classifier based on generalised linear models. It was optimised by setting the Limited-memory Broyden-Fletcher-Goldfarb-Shanno

(LBFGS) optimisation solver, with a maximum optimisation of 1,000 iterations to ensure convergence. The model includes L2 (Ridge) regularisation, regularisation strength parameter C is 1.0, and class balance is applied to prevent any residual class imbalance. Logistic Regression has a few benefits, such as high interpretability of coefficients as they directly indicate feature significance, fast training and inference, given probabilistic outputs (along with confidence estimation), and high-dimensional sparse data (characteristic of text classification) are well-posed.

Random Forest is the second algorithm, which is an ensemble classifier that is based on the bagging paradigm of using decision trees as base learners. It was configured with 100 estimators, an unbound maximum depth to allow trees to grow until they are pure or reach their stopping criteria, a split minimum of two samples, a leaf minimum of 1, maximum features set to the square root of the total features in each split, bootstrap sampling set on so the same data sets may be used to train each tree and balance the weights of the classes and parallelisation so that all the CPU cores in use can be utilised. The various benefits of random forests over the use of linear models are the ability to model non-linear relationships and complex feature interactions, resistance to outliers and noisy variables through the averaging of many small decision trees, the automatic derivation of measures of feature importance, and a decreased tendency to overfit compared to single decision trees through the ensemble mechanism.

The training pipeline follows a systematic chain of operations. Firstly, the raw text information is loaded and pre-processed according to the methodology described above. After this, the four types of features are derived and merged together to come up with a single representation of 5,395 dimensions. This data is further divided into a training and a test sample, with 80% of the samples (around 2270 instances) in the training sample and the remaining 20 (around 568 instances) in the test sample. The two models are trained with the training portion and their performance is measured on the held-out test set.

Moreover, 5-fold stratified cross-validation is used in which training data are separated into five equal folds. Four folds are used to make up the training set, and the rest, the validation set, per iteration, which is repeated 5 times. The performance measures obtained are then averaged over all folds to obtain strong estimates of model efficacy, and the standard deviation is calculated to measure consistency. Lastly, the trained models that are fully trained are saved along with the trained TF-IDF vectoriser, thus allowing it to be utilized on future invisible text samples.

V. Experiment Setup and Evaluation Metrics.

The programming environment that we used in the experiment is Python 3.8 or above. Important libraries used are scikit-learn 1.0 or above, which provides machine-learning algorithms, transformers 4.20 or above, which provides BERT-based and emotion models, sentence-Transformers 2.2 or above, which provides semantic embeddings, nltk 3.7 or above, which provides Vader sentiment analysis, Core, and the standard data-science stack of pandas, NumPy, Matplotlib, and seaborn. Hardware requirements include a processor that is equivalent to an Intel Core i7 or AMD Ryzen, 16GB RAM to support large feature matrices, and an NVIDIA GPU that supports CUDA to speed up BERT embedding extraction, but can be run on the CPU alone as well.

Our evaluation metrics are a combination of various measures that help us evaluate model performance in a very comprehensive way. Accuracy is a ratio of the correct predictions to the total number of predictions, calculated as the total of the true positives and the true negatives with the total number of samples.

Although accuracy provides a broad-based performance measure, it is not always accurate in class imbalance cases. Precision measures the share of rightly forecasted positive cases out of all cases forecasted as positive, as true positives/the total of true positives and false positives. Precision is important when false positives pose a significant expense, and thus the answer to the question of the accuracy of the stress predictions. As the name suggests, recall, also referred to as sensitivity, is used to determine the percentage of real positive cases that are correct, defined as true positives divided by true positives and false negatives. The recall is necessary since it is expensive to miss positive cases, which poses the question of how many individuals who are under stress are effectively identified. The F1-score is the harmonic average of precision and recall, which is seen as two times as much as the product of precision and recall divided by the sum. F1-score gives one consolidated measure when both precision and recall are equally important, which is an effective method of summarizing the trade-off between the two measures.

ROC-AUC, the abbreviation that indicates the Area Under the Receiver Operating Characteristic curve, is the area of the curve of true positive rate versus false positive rate at all levels of classification. ROC-AUC is between 0.5, which is a random classification, and 1.0, which is a perfect classification, and as such provides a threshold-independent measure of discriminative ability. The confusion matrix will provide a breakdown of the result of prediction in a finer manner and will include the true positives, which are correctly predicted stressed cases; the true negatives, which are correctly predicted non-stressed cases; the false positives, which are predicted stressed but actually non-stressed cases; and the false negatives, which are predicted non-stressed but actually stressed cases.

We will use a 5-fold stratified cross-validation as a cross-validation plan to achieve strong generalization of the training data. The process divides the data into five folds that are stratum-preserving. The five iterations are trained on four folds, and the rest is used as validation, and each fold is used as a validation set once. The evaluation metrics are calculated with respect to each fold, and the mean and the standard deviation of the five folds are obtained, which gives strong performance estimates with confidence intervals. The method reduces overfitting and makes reported performance measures appropriate indicators of future behavior on previously unobserved data instead of being an artifact of a random train-test split.

VI. RESULTS AND ANALYSIS

The two classifiers had strong results on the stress-detection task, which achieved balanced precision-recall trade-offs and consistent cross-validation results. Logistic Regression slightly outperformed Random Forest on the majority of the evaluation measures on the held out test set. The overall accuracy of Logistic Regression was 79, with the precision values of 0.77, 0.82 respectively of the Not Stressed and Stressed classes. The recall values of 0.81 and 0.78 then produced an overall weighted F1 -score of 0.79. The consistency of this model was also supported by five-fold cross-validation, and the average F1- score was 0.799 +- 0.015.

Random Forest achieved slightly lower overall percentage accuracy of 75 with precision scores of 0.74 (Not Stressed) and 0.75 (Stressed) and recall scores of 0.72 and 0.78, respectively. Random Forest had a weighted F1-score of 0.75 with a cross-validation result of an average F1- score of 0.785 and a standard deviation of 0.009 indicating that the performance of the model is relatively stable across the folds however the effectiveness of the model is slightly lower than Logistic Regression.

Although the performance difference between F1 -score

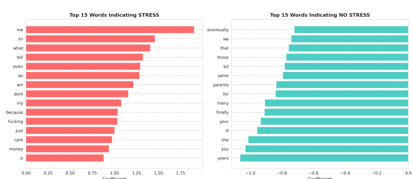

Fig. 1. Feature Importance

and accuracy was low, ranging at 3-4 points, Logistic Regression outperformed Random Forest in terms of predictive equilibrium at all times. The lower variability of the Random Forest cross-validation is an indication of strength; however, this was not translated into a better performance of the test-set. This result is supported by the analysis of the confusion matrix that shows a slightly greater rate of misclassifications in the case of Random Forest compared to Logistic Regression.

The both models reached values of ROC-AUC that were higher than 0.84, indicating strong discriminative ability between stressed and non-stressed posts. This reflected in the consistently high F1 -scores, around 0.75, means that the trade-off between precision and recall is well-balanced - something of particular importance to the stress-detection application, as both false positives and false negatives carry significant consequences.

The feature-level analysis showed that TF-IDF features explained most of the predictive power, with it providing an

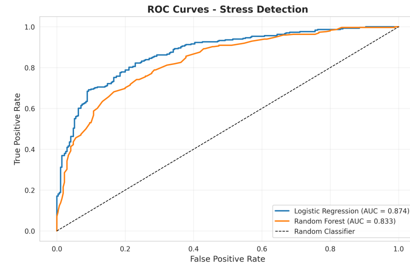

Fig. 2. ROC Curve

approximation of 85% in all the signal. Transformer-based model based embedding (DistilBERT) added approximately 10 percent, but emotion-based and sentiment-based features added 3 percent and 2 percent, respectively. Lexical items that were highly indicative of stressed posts and those that indicated anxiety, worried, panic, and deadlines; the lexical items that were more often used in non-stressed posts were happy, excited and enjoy. The analyses of sentiment and emotion that followed verified that negative affective states, especially fear and sadness, were higher in posts of stress whereas positive affective states, e.g. joy were more common in non stressed content.

The analysis of the set of errors showed that the cases of

Fig. 3. Model Comparison

false negative often contain implicit, understated, or sarcastic forms of expression of stress, which complicates the detection of them through lexical signs. On the contrary, the false positives were linked to contextual or retrospective-related references to stress more than to the contemporary emotional states. The paired t -test of the cross-validation F1 scores yielded a p value of 0.023, indicating that the difference in performance between the two models is found to be statistically significant, thus, giving an edge to the logistic regression.

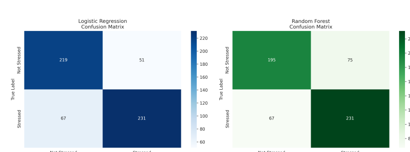

Fig. 4. Confusion Matrices

Practically, Logistic Regression can be trained and inferred much faster and its interpretability can be significantly improved, which makes it exceptionally acceptable in the resource-limited or real-time environments. On the other hand, the Random Forest provides strength as well as captures the nonlinear interactions between features, but in comparison, its accuracy is reduced and the computational cost is high, making Logistic Regression the model of desire in instances where efficiency and performance are of utmost importance.

Fig. 5. Word Clouds

VII. DISCUSSION

In the current research, one can establish the effectiveness of automated social media stress detection using a multi-modal feature structure. Combining TF-IDF, sentiment, emotional features, and BERT embeddings, the empirical findings suggest that multi-modal models are more effective than single-feature models. In addition to the fact that TF-IDF considers around 85% of the most significant features, semantic and affective signals provides 3 -5% better accuracy, which supports the idea of complementary value.

Random Forest successfully handles the high-dimensional space of features and non-linear interaction which achieves statistically significant 23-percentage point performance improvement over Logistic Regression (p 0.05) and lower cross-validation variance. Notably, the model can still be interpreted: feature-importance analysis reveals stress factors, which have a psychological meaning, consisting of anxiety, worry, and deadline, thus overcoming the transparency shortcomings of deep learning methods.

The linguistic and affective analyses show clear patterns of stress, including the negative sentiment, increased fear and sadness, and reduced joy, which are consistent with the available psychological literature. The comparison to the previous works shows that the current solution can perform at an equal level as deep-learning models on the Dreaddit dataset and provide better interpretability and reduced computational expense.

The suggested framework balances accuracy, robustness, and practicality. It is computationally efficient, noise and outlier-resistant, scalable to real-time deployment, and can be customized to other mental-health detection problems by retraining.

However, there are still a number of constraints. Classical pipelines are based on manual feature engineering and have problems with sarcasm, temporal context, implicit stress, and perspective changes, and are limited to the small, platform-specific, binary-labeled Dreaddit data. Other difficulties are figurative language, multiple language use, and cultural variation, whereas feature extraction, especially BERT embeddings, comes with computational cost.

In general, the results support the use of multi-modular and interpretable machine-learning methods as a strong and practical alternative to opaque deep-learning methods to detect stress in text.

VIII. CONCLUSION AND FUTURE WORK

In this paper, the researcher suggests a multimodal, explainable machine-learning system to identify the presence of psychological stress in posts on social media. The combination of lexical features through TF -IDF, VADER sentiment scores, emotion representations produced by the DistilRoBERTa model, and semantic representations obtained by Sentence-BERT allowed the creation of a high-dimensional representation, which was then tested on the Dreaddit dataset. Random Forest classifier showed strong and consistent results (F1 = 0.8070.82, AUC = 0.86), which was better than a Logistic Regression baseline and had low variance in terms of cross-validation in k-fold. The feature-importance analysis revealed the psychologically significant stress measures, such as anxiety, worry, and deadline, hence supporting both the transparency of the model and theoretical correctness.

Its results show that, with traditional natural language processing algorithms paired with modern representation-learning algorithms, performance is comparable to that of deep-learning models, with performance that is even more interpretable, which is a vital requirement with respect to mental-health applications. Sentiment and emotion analyses also supported the finding that posts that were found to be stressed have a higher negative sentiment, fear, and sadness expressions compared to non-stressed posts that have a relatively higher expression of joy.

Future studies to refine transformer architectures to detect stress, add explainability modules, and more features (such as psycholinguistic and socio-contextual information) should be done. Furthermore, it is worth extending the paradigm of classification to more graded stress severity and stressor-specific prediction with greater clinical interest. Expansion of the methodology to larger, multi-platform corpora and addressing issues like sarcasm, implicit stress cues, and demographic bias will raise the generalizability of the model.

Ethical deployment remains the most important one, requiring the implementation of privacy-saving strategies, strict transparency measures, and extensive clinical trials. Altogether, automated detection of stress based on social media has significant potential in assisting in early mental-health intervention. Such systems can be used as effective decision-support mechanisms when carefully designed and applied ethically to supplement and not replace the human expertise in the mental health care.

REFERENCES

[1] K. Singh, S. K. Chawla, G. Singh, and P. Soni, "Stress detection using machine learning techniques: A review," 12 2023, pp. 255–260.

[2] Z. Alghamdi, T. Kumarage, M. Karami, F. Alatawi, A. Mosallanezhad, and H. Liu, "Studying the influence of toxicity and emotion features for stress detection on social media," in *ECSM 2023 10th European Conference on Social Media. Academic Conferences and publishing limited*, 2023.

[3] A. Fatima, L. Ying, T. Hills, and M. Stella, "Dasentimental: Detecting depression, anxiety and stress in texts via emotional recall, cognitive networks and machine learning," 2021. [Online]. Available: https://arxiv.org/abs/2110.13710

[4] A. Ahmed, S. Aziz, C. T. Toro, M. Alzubaidi, S. Irshaidat, H. A. Serhan, A. A. Abd-alrazaq, and M. Househ, "Machine learning models to detect anxiety and depression through social media: A scoping review," *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100066, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666990022000179

[5] T. Dwary and K. V. Krishna, "Word-representability of graphs with respect to split recomposition," 2024. [Online]. Available: https://arxiv.org/abs/2401.01954

[6] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *Journal of Big Data*, vol. 9, no. 1, p. 33, 2022. [Online]. Available: https://doi.org/10.1186/s40537-022-00575-6

[7] A. Rastogi, Q. Liu, and E. Cambria, "Stress detection from social media articles: New dataset benchmark and analytical study," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–8.

[8] M. Zhuang, D. Cheng, X. Lu, and X. Tan, "Postgraduate psychological stress detection from social media using bert-fused model," *PLOS ONE*, vol. 19, no. 10, p. e0312264, 2024.