Stress Identification in Online Social Networks

Ashok Kumar, Tina Esther Trueman
Department of Information Science and Technology
Anna University
Chennai, India
{jashokkumar83, tina_trueman}@auist.net

Erik Cambria
School of Computer Science and Engineering
Nanyang Technological University
Singapore
cambria@ntu.edu.sg

***Abstract*—Online social networks have become one of the primary ways of communication to individuals. It rapidly generates a large volume of textual and non-textual data such as images, audio, and videos. In particular, textual data plays a vital role in detecting mental health-related problems such as stress, depression, anxiety, and emotional and behavioral disorders. In this paper, we identify the mental stress of online users in social networks using a transformers-based RoBERTa model and an autoregressive model, also called XLNet. We implement this model in both a constrained system and an unconstrained system. The constrained system maintains the gold standard datasets such as training, validation, and testing. On the other hand, the unconstrained system divides the given dataset into user-specific training, validation, and test sets. Our results indicate that the proposed transformers-based RoBERTa model achieves a better result in both constrained and unconstrained systems than the state-of-the-art models.**

***Index Terms*—Stress detection, Mental health, Transformers**

I. Introduction

Nowadays, online social media has transformed the way of communication or interaction between people. They share their factual information, day-to-day activities, feelings, opinions, experiences, desires, hopes, and emotions in the form of tweets, posts, and messages [1]. This textual information can be used to identify mental illness in individuals [2]–[6]. Psychological stress is an emotional feeling or a physical tension of an individual. In real life, everyone experiences psychological stress from time to time due to any thought or event that makes them feel angry, nervous, or frustrated [7].

Stress is broadly studied into two types, namely, acute stress and chronic stress [8]. Firstly, acute stress is a short-term or temporary condition of an individual that goes away quickly. For instance, doing something new, fight with a partner, loss of a loved one, the threat of death, natural disasters, sexual assault, or motor vehicle accidents. Sometimes, it helps an individual to manage dangerous situations like meeting a deadline or learning a skill. Secondly, chronic stress is a long-lasting feeling of an individual that severely damages human body reactions such as irritability, difficulty sleeping, digestive problems, feeling helpless, nervousness, headaches, difficulty concentrating, high blood pressure, diabetes, heart disease, and skin problems.

Chronic stress can come from various thoughts such as an unhappy married life, high-pressure jobs, money problems, or trouble at work [9]. Therefore, stress detection has become an important research field in online social networks such as Facebook, Twitter, and Reddit.

For instance, here is an example of how a social media user expresses stress: “Hi, I am a 19-year-old kid dealing with Agoraphobia over the past year. Very boldly said, Agoraphobia is the fear of going to public places, sometimes even leave the house. The past week, some very unpleasant things have happened in my household. I told my parents about what I have been dealing with a while back" [10]. Traditionally, researchers studied the psychological stress of individuals using a questionnaire, face-to-face interview, and wearable sensors. However, these traditional methods are time-costing, labor-consuming, and hysteric [11]. Recently, researchers developed automatic stress detection in online social networks using machine learning and deep learning architectures. The machine learning architectures or models use hand-crafted feature engineering techniques to identify stress. Similarly, deep learning architectures use semantic context vectors to identify stress in online social networks. More recently, transformers-based [12] approaches such as encoder architecture (BERT, RoBERTa) [13], [14] and decoder architecture (XLNet) [15] gained popularity among researchers in the field of natural language inference, sentiment analysis, question answering, and recommendation systems [16]–[18].

In this paper, we use a RoBERTa and an XLNet-based classifier for stress detection in users post. The RoBERTa is defined with the encoder architecture of the transformer with an architecture similar to BERT. Similarly, the XLNet is defined with the decoder architecture of the transformer with a similar architecture of BERT. Particularly, these pre-trained models outperform well than BERT-based classifiers. It is also very effective to deal with unstructured data and to learn long-textual contextual information. The rest of this paper is organized as follows: Section 2 illustrates existing research works in stress detection; Section 3 explains transformer-based stress identification in online social networks; Section 4 presents the results and discussion; finally, Section 5 proposes concluding remarks.

II. Related works

In the last decade, online social media has become one of the most important platforms for people to share their mental health issues. Researchers use this information and develop new techniques using natural language processing (NLP) and machine learning to assist people. In this section, we briefly describe the existing works in stress detection.

Murarka et al. [3] applied RoBERTa architecture to detect and classify online mental health illness posts into ADHD, anxiety, bipolar disorder, depression, and PTSD. Their results indicated that the proposed RoBERTa model achieves 86%, 72%, and 89% F1-scores for posts, titles, and posts and titles, respectively. Lin et al. [11] proposed a hybrid model that combines a graph model with convolutional neural networks (CNN) for the psychological stress detection task using users' social interactions. The authors used textual, visual, and social features for stress detection. They improved the model performance by a 6-9% F1-score.

Iranfar et al. [19] detected stress from multimodal signals using a machine learning framework, which includes feature identification, outlier detection, imputation, and classification. The authors effectively addressed missing data and outliers. Their study indicated that the eXtreme Gradient Boosting (XGB) algorithm achieves a higher cross-validation score. Hosseini et al. [20] provided a biometric stress detection dataset that is collected from nurses during the COVID-19 pandemic. It contains stress events, survey responses, and signals. The authors have captured the psychological data and its associated events. Also, they performed ANOVA and Tukey's test to identify the differences between stress groups.

Li [21] developed a stress detection system with an early warning using wireless network transmission. The authors implemented this system in a real-time hardware and software setup. They used a predefined psychological stress index threshold to measure the early warning. Dacunhasilva et al. [22] detected stress with pressure sensors using instrumented peripherals such as keyboards and mice. They validated this system design in the laboratory environment by comparing conventional features (mouse trajectories, keystroke dynamics) and augmented features from pressure sensors.

Gonzalez-Carabarin et al. [23] proposed a multi-sensing system to evaluate personalized stress levels based on EEG-ECG markers. The authors used 24 individuals between the age of 18-23. They used semisupervised machine learning models to process acquired signals and five supervised machine learning models (KNN, SVM, RT, FT, and ANN) to categorize the obtained clusters. Their study indicated that SVM achieves 79.91% test accuracy. Mou et al. [24] developed a framework to detect driver stress using multimodal fusion. They employed an attention-based CNN-LSTM model to fuse vehicle data, eye data, and environmental data. The authors achieved 95.5% accuracy on driver stress detection.

Also, Bara et al. [25] used the GRU model for multimodal stress detection. The authors evaluated this task using different signals. Their study indicated that a 1.3% improvement in close-up video and 2.4% in physiological sensors. Delmastro et al. [26] detected stress using wearable sensors and machine learning from mild cognitive impairment frail older adults. The authors were applied correlation, information gain, and principal component analysis features for the psychological stress input data. Then, they employed machine learning algorithms. Their study indicated that the Random Forest and AdaBoost algorithms outperform the other algorithms.

Dham et al. [27] automated the mental stress detection based on stacking classifier and RBF (radial basis function). They used wearable sensor data that is collected from stressed individuals. Li et al. [28] proposed an algorithm based on correlation learning and clinical patient questionnaire-based lexicon to detect stress symptoms related to COVID-19 at a spatiotemporal scale. This algorithm addresses the limitation of topic modeling and ambiguity minimization. Their study suggested a strong relationship between increased COVID-19 cases and stress symptoms in the major cities of the United States.

Albertetti et al. [29] proposed deep learning approaches to detect stress using psychological signals in a laboratory condition. The authors used a decision tree (DT), recurrent neural networks (RNN), and Convolutional RNN for the binary-level stress detection tasks. They achieved the highest accuracy of 71% using RNN. Gopalakrishna Pillai et al. [30] developed a lexicon-based TensiStrength algorithm to detect stress and relaxation by incorporating word sense disambiguation (WSD). Their study indicated that the proposed TensiStrength algorithm with WSD achieves better results than machine learning algorithms. In summary, the existing researchers used hand-crafted features, context-independent features, and lexicon-based features to identify stress in a given text. Thus, we present a generalized autoregressive model (XLNet) with long-context-dependent features to detect stress in online social networks.

III. Stress Identification

We present a transformers-based stress identification model in online social networks. Specifically, we use the RoBERTa model and the generalized autoregressive model, also called XLNet for the task of stress identification as in Fig. 1. We explain the proposed model in detail in the following steps.

A. Dataset

We use a Reddit dataset for stress identification in online social networks [10]. This dataset contains users' posts in five domains with ten subreddits, namely, abuse: domestic violence and survivors of abuse, anxiety: anxiety and stress, financial: almost homeless, assistance, food pantry, and homeless, Post-Traumatic Stress Disorder: PTSD and relationships, and social: relationships. Overall, the dataset contains 3553 labeled posts (stress or non-stress) and their LIWC (Linguistic Inquiry and Word Count) features, syntactic features, and social media features. It is further divided into a training set and a test set. They contain 2838 and 715 users' posts with LIWC features, respectively.

B. Preprocessing

We apply mainly three preprocessing techniques to the given training and test sets. Firstly, we fix the broken Unicode in the given training sets and test sets using FTFY (fixes text for you) library [31]. Secondly, we apply a contraction map technique to expand the shortened English tokens like "im" into "I am", "imma" into "I am going to" and so on.

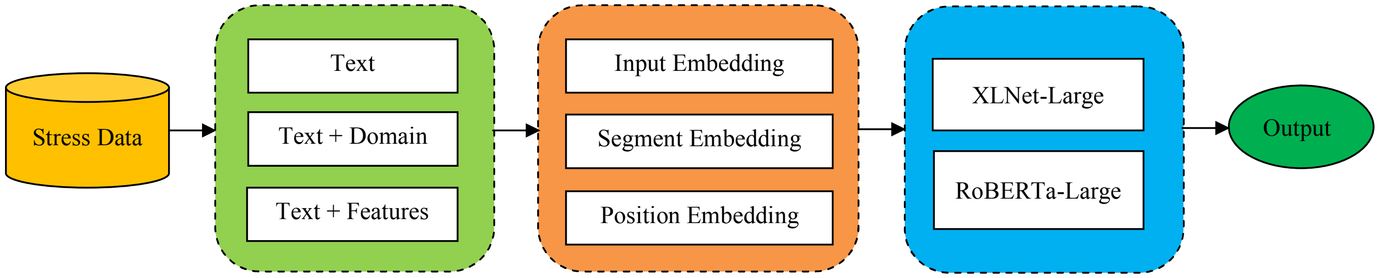


Fig. 1. The proposed stress detection model

Thirdly, we use Sentic Parser [32] to perform other syntactic and semantic preprocessing tasks such as negation handling, microtext normalization, and compound word processing. The parser removes inflections, such as *-ing*, *-ful*, and *-able*, and neutral prefixes, such as *en-*, *re-*, and *co-* (Fig. 2), so that words like 'entrust', 'entrusts', 'entrusted', 'entrusting', 'entrustment', 'entrustments', 'trustable', 'trustability', 'trusts', 'trusted', 'trustful', 'trustfully', 'trustfulness', 'trustily', 'trustiness', 'trusting', 'trustingly', 'trustingness', 'trustworthy', 'trustworthily', 'trustworthiness', etc. are all normalized to `trust`. The same mechanism applies to other non-canonical suffixes such as *-like*, *-hood*, *-dom*, and *-ship* so that words like 'saintlike', 'sainthood', 'saintdom', and 'saintship' are all normalized as `saint`.

Sentic Parser also handles negative prefixes such as *mis-*, *dis-*, and *un-* so that words like 'distrust' and 'mistrust' can be normalized as `NOT trust`. Such negative prefix handling happens concomitantly with inflection removal so that also words like 'distrusts', 'distrusted', 'distrustable', 'distrustful', 'distrustfully', 'distrustfulness', 'distrusting', 'distrustingly', 'mistrusts', 'mistrusted', 'mistrustable', 'mistrustful', 'mistrustfully', 'mistrustfulness', 'mistrusting', 'mistrustingly', 'trustless', 'untrustworthy', 'untrusty', 'untrusting', and more, are all normalized as `NOT trust`.

Thanks to such a mechanism, which leverages both inflectional and derivational morphology, Sentic Parser is also able to decode wrong English expressions such as 'stucked', 'accessable', or 'inglamorous', which can be rather common in social media text. The same rule of checking whether a concept is present in SenticNet still applies here so that concepts like `defraud`, `distress`, and `disclose` are not wrongly normalized as `NOT fraud`, `NOT stress`, and `NOT close`, respectively. Finally, Sentic Parser also performs microtext normalization [33] so that words like 'gooooooood' and 'horrrible' are normalized as 'good' and 'horrible', respectively.

C. Text representation

We first represent the preprocessed text with special tokens in the format of $[CLS]$ $S1$ $[SEP]$ $S2$ $[SEP]$, where $[CLS]$ and $[SEP]$ indicate special tokens at the beginning of the sequence and end of the sequence, and $S1$ and $S2$ indicate multiple input sequences [12], [13], [15].

Second, the input sequences are tokenized into subwords and characters. It is then represented into token, segment, and attention embeddings. Finally, we apply these embeddings into the pre-trained RoBERTa and XLNet models, two language models similar to BERT but with different training corpus and language modeling objective. These models provide long-context-dependent information of the given input tokens and have been proven very effective for affective computing and sentiment analysis [34]–[37].

D. XLNet-Based Classifier

XLNet is one of the latest techniques in the field of natural language processing [15], [38]. It performs better than BERT [13] and overcomes the limitations of BERT. It is used in various tasks such as sentiment analysis, question-answering, text classification, etc. The BERT model tries to discover the masked words in a sentence, but XLNet works differently. It is built on the previous state-of-the-art methods. XLNet uses the best of both auto-encoder language models as used in BERT and auto-regressive language models as used in GPT-2.

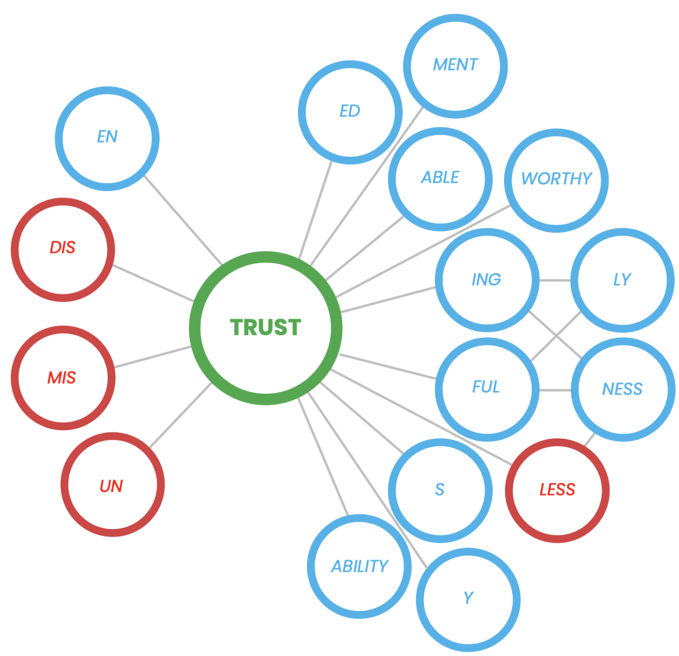


Fig. 2. A sample parse graph for word inflections.

Methods	Class	Validation		Testing	
		NS	S	NS	S
Text	NS	114	21	274	72
	S	27	122	54	315
Text + Domain	NS	91	44	239	107
	S	11	138	35	334
Text + Features	NS	104	31	255	91
	S	12	137	41	328

TABLE I
CONFUSION MATRIX FOR THE CONSTRAINED SYSTEM USING XLNET

Methods	Class	Validation		Testing	
		NS	S	NS	S
Text	NS	127	26	135	35
	S	17	150	23	163
Text + Domain	NS	125	28	140	30
	S	28	139	37	149
Text + Features	NS	117	36	122	48
	S	14	153	18	168

TABLE II
CONFUSION MATRIX FOR THE UNCONSTRAINED SYSTEM USING XLNET

Methods	Class	Validation		Testing	
		NS	S	NS	S
Text	NS	110	25	278	68
	S	23	126	52	317
Text + Domain	NS	102	33	255	91
	S	17	132	39	330
Text + Features	NS	113	22	280	66
	S	19	130	50	319

TABLE III
CONFUSION MATRIX FOR THE CONSTRAINED SYSTEM USING ROBERTA

Methods	Class	Validation		Testing	
		NS	S	NS	S
Text	NS	129	24	136	34
	S	30	137	23	163
Text + Domain	NS	119	34	126	44
	S	15	152	17	169
Text + Features	NS	113	40	119	51
	S	10	157	14	172

TABLE IV
CONFUSION MATRIX FOR THE UNCONSTRAINED SYSTEM USING ROBERTA

The auto-encoder language model can work in both directions, whereas the auto-regressive language model can works either in forward or reverse directions. Auto-regressive language models are good at the text generation tasks, whereas auto-encoder language models are good at seeing the context in both directions. Therefore, XLNet uses the auto-regressive language model with a bidirectional context via factorization order. This model is developed based on the decoder-architecture of the transformer model [12]. In particular, XLNet takes a list of tokens, where each token is some word and outputs the probability of occurrence of some word in the list. The model has enough information that it can predict what comes next from the input sequence words via factorization order. Let $T = \{He, likes, this, shirt\}$ be the given input text. XLNet uses the permutation order of this given input sequence is 4! For instance, we consider two input sequence order $[this \to he \to likes \to shirt]$ and $[He \to likes \to shirt \to this]$.

We then predict the word 'this' in both sequences. The first sequence returns the probability of this word as $P(this)$, whereas the input sequence does not have preceding words to consider. Similarly, the second sequence considers both preceding and succeeding words as $P(this \mid He, likes, shirt)$. More specifically, the XLNet model is trained on a corpus of 30 billion words with two variants, namely, the XLNet-Base model and the XLNet-Large model. These pre-trained models use permutations, attention masks, and two-stream self-attentions, namely, content stream and query stream self-attentions. These greatly help to achieve better performance. In this work, we detect stress in online social networks using the XLNet-Large fine-tuning model. This fine-tuning model is designed with 340M parameters based on 24 decoder layers, 16 attention heads, and 1024 hidden representations.

E. RoBERTa-Based Classifier

RoBERTa is also one of the latest techniques in the field of natural language processing [14]. It is the modified version of the BERT model for improving the performance of downstream tasks. Specifically, RoBERTa addresses four training procedures; dynamic masking, model input format and NSP (next sentence prediction), training with large mini-batches, and text encoding. Firstly, the BERT model uses only a single static masking pattern during data preprocessing and training. In contrast, RoBERTa uses a dynamic masking pattern for an input sequence when it feeds to the model every time. Secondly, the BERT model observes concatenated document segments either from the same document or distinct documents through the NSP loss. It trains the model with the input format of SEGMENT-PAIR+NSP loss. The RoBERTa model uses FULL-SENTENCES without NSP loss for improving the performance of downstream tasks. Thirdly, the original BERT model trains 256 input sequences in a batch. In contrast, RoBERTa uses increased batch sizes (2K and 8K) for training the model. Finally, the RoBERTa model uses character-level Byte-Pair Encoding (BPE) with increased vocabulary size (50K). The BERT model uses only a vocabulary size of 30K.

These modified training procedures improve the performance of downstream tasks than the original BERT. More specifically, RoBERTa has trained on five corpora over 160GB texts with two variants, namely, the RoBERTa-Base model and the RoBERTa-Large model. These pre-trained models use self-attentions (A) and FNN (feed-forward neural network). These greatly help the transformer models to achieve better performance in downstream tasks. In this work, we detect stress in online social networks using the RoBERTa-Large fine-tuning model. This fine-tuning model is designed with 355M parameters based on 24 encoder layers, 16 attention heads, and 1024 hidden representations.

IV. RESULTS AND DISCUSSION

We implement the stress identification task in online social networks in a constrained system and unconstrained systems using the RoBERTa-Large and XLNet-Large model. The constrained system uses the given training set and testing set without any modifications or changes.

Methods	Class	Validation			Testing		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Text	Non Stress	0.8085	0.8444	0.8261	0.8354	0.7919	0.8131
	Stress	0.8531	0.8188	0.8356	0.8140	0.8537	0.8333
	MacF1	0.8308	0.8316	0.8309	0.8247	0.8228	0.8232
	MicF1	0.8310	0.8310	0.8310	0.8238	0.8238	0.8238
Text + Domain	Non Stress	0.8922	0.6741	0.7679	0.8723	0.6908	0.7710
	Stress	0.7582	0.9262	0.8338	0.7574	0.9051	0.8247
	MacF1	0.8252	0.8001	0.8009	0.8148	0.7980	0.7978
	MicF1	0.8063	0.8063	0.8063	0.8014	0.8014	0.8014
Text + Features	Non Stress	0.8966	0.7704	0.8287	0.8615	0.7370	0.7944
	Stress	0.8155	0.9195	0.8644	0.7828	0.8889	0.8325
	MacF1	0.8560	0.8449	0.8465	0.8222	0.8129	0.8134
	MicF1	0.8486	0.8486	0.8486	0.8154	0.8154	0.8154

TABLE V
PERFORMANCE OF THE CONSTRAINED SYSTEM USING XLNET-LARGE

Methods	Class	Validation			Testing		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Text	Non Stress	0.8819	0.8301	0.8552	0.8544	0.7941	0.8232
	Stress	0.8523	0.8982	0.8746	0.8232	0.8763	0.8490
	MacF1	0.8671	0.8641	0.8649	0.8388	0.8352	0.8361
	MicF1	0.8656	0.8656	0.8656	0.8371	0.8371	0.8371
Text + Domain	Non Stress	0.8170	0.8170	0.8170	0.7910	0.8235	0.8069
	Stress	0.8323	0.8323	0.8323	0.8324	0.8011	0.8164
	MacF1	0.8247	0.8247	0.8247	0.8117	0.8123	0.8117
	MicF1	0.8250	0.8250	0.8250	0.8118	0.8118	0.8118
Text + Features	Non Stress	0.8931	0.7647	0.8239	0.8714	0.7176	0.7871
	Stress	0.8095	0.9162	0.8596	0.7778	0.9032	0.8358
	MacF1	0.8513	0.8404	0.8417	0.8246	0.8104	0.8115
	MicF1	0.8438	0.8438	0.8438	0.8146	0.8146	0.8146

TABLE VI
PERFORMANCE OF THE UNCONSTRAINED SYSTEM USING XLNET-LARGE

Methods	Class	Validation			Testing		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Text	Non Stress	0.8271	0.8148	0.8209	0.8424	0.8035	0.8225
	Stress	0.8344	0.8456	0.8400	0.8234	0.8591	0.8408
	MacF1	0.8308	0.8302	0.8304	0.8329	0.8313	0.8317
	MicF1	0.8310	0.8310	0.8310	0.8322	0.8322	0.8322
Text + Domain	Non Stress	0.8571	0.7556	0.8031	0.8673	0.7370	0.7969
	Stress	0.8000	0.8859	0.8408	0.7838	0.8943	0.8354
	MacF1	0.8286	0.8207	0.8220	0.8256	0.8157	0.8162
	MicF1	0.8239	0.8239	0.8239	0.8182	0.8182	0.8182
Text + Features	Non Stress	0.8561	0.8370	0.8464	0.8485	0.8092	0.8284
	Stress	0.8553	0.8725	0.8638	0.8286	0.8645	0.8462
	MacF1	0.8557	0.8548	0.8551	0.8385	0.8369	0.8373
	MicF1	0.8556	0.8556	0.8556	0.8378	0.8378	0.8378

TABLE VII
PERFORMANCE OF THE CONSTRAINED SYSTEM USING ROBERTA-LARGE

In the unconstrained system, we combine both training and testing sets. We then randomly split the entire dataset into 80% for the training set, 10% for the validation set, and 10% for the testing set. Specifically, we use the Reddit stress detection dataset that contains a training set of 2838 posts and their related LIWC, syntactic, and social media features, and a test set of 715 posts and their related LIWC, syntactic, and social media features. We use the following baselines models for online users stress detection, namely, Logistic Regression (LR) with Word2Vec, LR with BERT, CNN, and CNN with attention, Gated RNN (GRNN), GRNN with attention, and BERT. For the constrained system, 10% percent of user posts were taken for validation from the original training set.

Now, the training set contains 2554 user posts, and the validation set contains 284 user posts. Similarly, the unconstrained system divides the entire dataset into 2877 user posts for the training set, 320 user posts for the validation set, and 356 user posts for the test set. In this work, we represented three types of features, namely, text features, text and domain features which include all subreddits, and text and their related features such as LIWC, syntactic, and social media features. We then performed XLNet large model on these feature representations in both constrained and unconstrained systems using Google Colab Pro. The constrained system uses the following hyperparameters: 310 sequence length (310), batch sizes (6), learning rate (2e-5), and epochs (3).

Methods	Class	Validation			Testing		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Text	Non Stress	0.8113	0.8431	0.8269	0.8553	0.8000	0.8267
	Stress	0.8509	0.8204	0.8354	0.8274	0.8763	0.8512
	MacF1	0.8311	0.8317	0.8311	0.8414	0.8382	0.8390
	MicF1	0.8313	0.8313	0.8313	0.8399	0.8399	0.8399
Text + Domain	Non Stress	0.8881	0.7778	0.8293	0.8811	0.7412	0.8051
	Stress	0.8172	0.9102	0.8612	0.7934	0.9086	0.8471
	MacF1	0.8526	0.8440	0.8452	0.8373	0.8249	0.8261
	MicF1	0.8469	0.8469	0.8469	0.8287	0.8287	0.8287
Text + Features	Non Stress	0.9187	0.7386	0.8188	0.8947	0.7000	0.7855
	Stress	0.7970	0.9401	0.8626	0.7713	0.9247	0.8411
	MacF1	0.8578	0.8393	0.8407	0.8330	0.8124	0.8133
	MicF1	0.8438	0.8438	0.8438	0.8174	0.8174	0.8174

TABLE VIII
PERFORMANCE OF THE UNCONSTRAINED SYSTEM USING ROBERTA-LARGE

Methods	Accuracy
Majority baseline	68.08
CNN	71.82
CNN + Features	70.35
GRNN_Attn	73.55
GRNN_Attn + Features	72.86
N-gram baseline	74.41
N-gram + Features	77.00
Logreg_Word2Vec + Features	77.06
Logreg_BERT + Features	79.37
Logreg_Domain Word2Vec + Features	79.80
BERT-base	80.65
Proposed XLNet + Text (C)	82.38
Proposed RoBERTa + Text (C)	83.22
Proposed XLNet + Text (U)	83.71
Proposed RoBERTa + Text (U)	83.99
Proposed XLNet + Text + Domain (C)	80.14
Proposed RoBERTa + Text + Domain (C)	81.82
Proposed XLNet + Text + Domain (U)	81.18
Proposed RoBERTa + Text + Domain (U)	82.87
Proposed XLNet + Text + Features (C)	81.54
Proposed RoBERTa + Text + Features (C)	83.78
Proposed XLNet + Text + Features (U)	81.46
Proposed RoBERTa + Text + Features (U)	81.74

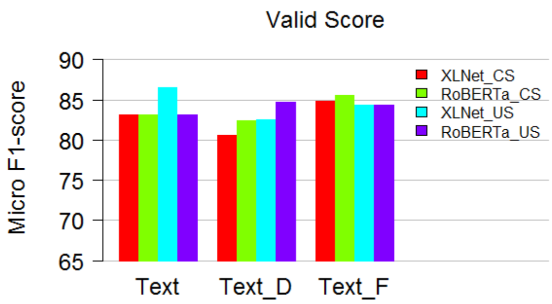
TABLE IX
RESULT COMPARISON

Similarly, the unconstrained system uses the following hyperparameters: sequence length (310), batch sizes (4), learning rate (2e-5), epochs (6), and steps (4320). Table I and Table II show the validation and testing confusion matrix for text, text with domain, and text with features (LIWC, syntactic, and social media) for the XLNet model with constrained and unconstrained systems. Table III and Table IV show the validation and testing confusion matrix for the RoBERTa model with constrained and unconstrained systems.

Table V and Table VI show the XLNet-based constrained and unconstrained systems' performance with precision, recall, and F1-score and their micro and macro scores [39] for all feature representations. These tables indicate that the XLNet large model achieves the micro F1-score of 83.10% and 86.56% for the validation set and 82.38% and 83.71% for the test set with text feature representation. Also, the model achieves an 80.63% and 82.50% micro F1-score for the validation set and 80.14% and 81.18% micro F1-scores for the validation set with text and domain feature representations.

As for the text and feature representations, the model achieves 84.86% and 84.38% micro F1-scores for the validation set and 81.54% and 81.46% for the test set. Moreover, Table VII and Table VIII show the RoBERTa-based constrained and unconstrained systems' performance with precision, recall, and F1-score and their micro and macro scores for all feature representations. These tables indicate that the RoBERTa large model achieves the micro F1-score of 83.10% and 83.13% for the validation set and 83.22% and 83.99% for the test set with the text feature representation.

Also, the model achieves an 82.39% and 84.69% micro F1-score for the validation set and 81.82% and 82.87% micro F1-scores for the test set with text and domain feature representations. Similarly, for the text and feature representations, the model achieves 85.56% and 84.38% micro F1-scores for the validation set and 83.78% and 81.74% for the test set. These results are visualized in Fig. 2. Result comparison of the proposed RoBERTa and XLNet large model with baseline models is shown in Table IX.



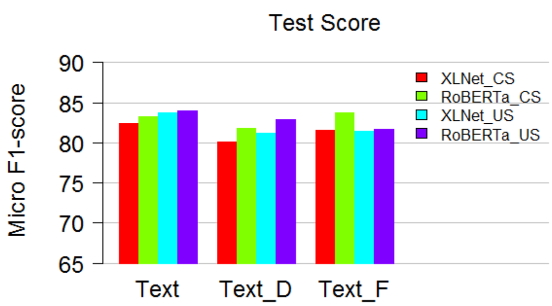


Fig. 3. Visualization of the validation and test scores

This table indicates that the proposed model achieves higher accuracy for text feature representation and text with LIWC, syntactic, and social media features in the constrained and unconstrained systems. Furthermore, the unconstrained system performs well than the constrained system for text feature representation and text and features representations. Overall, the RoBERTa model performs well both constrained and unconstrained systems.

V. Conclusion

In this paper, we presented stress identification in online social networks using a transformers-based XLNet model. The proposed XLNet model is implemented in the constrained system and unconstrained systems. Specifically, we represented three types of features such as text features, text and domain features, and text and their related features such as LIWC, syntactic, and social media features. Our results reveal that the proposed transformers-based XLNet model achieves better results in both constrained and unconstrained systems than the existing state-of-the-art models. As future work, we also plan to detect stressed users based on gender and age group in online social networks as we have done in our previous work on sentiment detection [40].

VI. Acknowledgements

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its AME Programmatic Funding Scheme (Project #A18A2b0046).

References

[1] D. Camacho, A. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Information Fusion*, vol. 63, pp. 88–120, 2020.

[2] E. Cambria, T. Benson, C. Eckl, and A. Hussain, "Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 533–10 543, 2012.

[3] A. Murarka, B. Radhakrishnan, and S. Ravichandran, "Detection and classification of mental illnesses on social media using roberta," *arXiv preprint arXiv:2011.11226*, 2020.

[4] S. Ji, X. Li, Z. Huang, and E. Cambria, "Suicidal ideation and mental disorder detection with attentive relation networks," *Neural Computing and Applications*, vol. 34, pp. 10 309–10 319, 2022.

[5] S. Ji, T. Zhang, L. Ansari, J. Fu, P. Tiwari, and E. Cambria, "MentalBERT: Publicly available pretrained language models for mental healthcare," in *LREC*, 2022, pp. 7184–7190.

[6] E. Cambria, A. Hussain, T. Durrani, C. Havasi, C. Eckl, and J. Munro, "Sentic computing for patient centered applications," in *IEEE ICSP*, 2010, pp. 1279–1282.

[7] Y. Susanto, A. Livingstone, B. C. Ng, and E. Cambria, "The hourglass model revisited," *IEEE Intelligent Systems*, vol. 35, no. 5, pp. 96–102, 2020.

[8] MedlinePlus, 2021. [Online]. Available: https://medlineplus.gov/ency/article/003211.htm

[9] A. Kandola, "What are the health effects of chronic stress?" 2018. [Online]. Available: https://www.medicalnewstoday.com/articles/323324#signs-and-symptoms

[10] E. Turcan and K. McKeown, "Dreaddit: A reddit dataset for stress analysis in social media," in *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, 2019, pp. 97–107.

[11] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T.-S. Chua, "Detecting stress based on social interactions in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1820–1833, 2017.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT (1)*, 2019.

[14] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[15] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Advances in neural information processing systems*, vol. 32, 2019.

[16] E. Cambria, B. Schuller, B. Liu, H. Wang, and C. Havasi, "Statistical approaches to concept-level sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 6–9, 2013.

[17] I. Chaturvedi, Y.-S. Ong, I. Tsang, R. Welsch, and E. Cambria, "Learning word dependencies in text by means of a deep recurrent belief network," *Knowledge-Based Systems*, vol. 108, pp. 144–154, 2016.

[18] N. Howard and E. Cambria, "Intention awareness: Improving upon situation awareness in human-centric environments," *Human-centric Computing and Information Sciences*, vol. 3, no. 9, 2013.

[19] A. Iranfar, A. Arza Valdes, and D. Atienza Alonso, "Relearn: A robust machine learning framework in presence of missing data for multimodal stress detection from physiological signals," in *43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, no. CONF, 2021.

[20] S. Hosseini, S. Katragadda, R. T. Bhupatiraju, Z. Ashkar, C. W. Borst, K. Cochran, and R. Gottumukkala, "A multi-modal sensor dataset for continuous stress detection of nurses in a hospital," *arXiv preprint arXiv:2108.07689*, 2021.

[21] Y. Li, "Psychological stress detection and early warning system based on wireless network transmission," *Scientific Programming*, vol. 2021, 2021.

[22] D. R. Dacunhasilva, Z. Wang, and R. Gutierrez-Osuna, "Towards participant-independent stress detection using instrumented peripherals," *IEEE Transactions on Affective Computing*, 2021.

[23] L. Gonzalez-Carabarin, E. Castellanos-Alvarado, P. Castro-Garcia, and M. Garcia-Ramirez, "Machine learning for personalised stress detection: Inter-individual variability of eeg-ecg markers for acute-stress response," *Computer Methods and Programs in Biomedicine*, vol. 209, p. 106314, 2021.

[24] L. Mou, C. Zhou, P. Zhao, B. Nakisa, M. N. Rastgoo, R. Jain, and W. Gao, "Driver stress detection via multimodal fusion using attention-based cnn-lstm," *Expert Systems with Applications*, vol. 173, p. 114693, 2021.

[25] C.-P. Bara, M. Papakostas, and R. Mihalcea, "A deep learning approach towards multimodal stress detection." in *AffCon@ AAAI*, 2020, pp. 67–81.

[26] F. Delmastro, F. Di Martino, and C. Dolciotti, "Cognitive training and stress detection in mci frail older people through wearable sensors and machine learning," *IEEE Access*, vol. 8, pp. 65 573–65 590, 2020.

[27] V. Dham, K. Rai, and U. Soni, "Mental stress detection using artificial intelligence models," in *Journal of Physics: Conference Series*, vol. 1950, no. 1. IOP Publishing, 2021, p. 012047.

[28] D. Li, H. Chaudhary, and Z. Zhang, "Modeling spatiotemporal pattern of depressive symptoms caused by covid-19 using social media data mining," *International Journal of Environmental Research and Public Health*, vol. 17, no. 14, p. 4988, 2020.

[29] F. Albertetti, A. Simalastar, and A. Rizzotti-Kaddouri, "Stress detection with deep learning approaches using physiological signals," in *International Conference on IoT Technologies for HealthCare*. Springer, 2020, pp. 95–111.

[30] R. Gopalakrishna Pillai, M. Thelwall, and C. Orasan, "Detection of stress and relaxation magnitudes for tweets," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 1677–1684.

[31] R. Speer, "ftfy," Zenodo, 2019, version 5.5. [Online]. Available: https://doi.org/10.5281/zenodo.2591652

[32] E. Cambria, R. Mao, S. Han, and Q. Liu, "Sentic parser: A graph-based approach to concept extraction for sentiment analysis," in *Proceedings of ICDM Workshops*, 2022.

[33] R. Satapathy, E. Cambria, A. Nanetti, and A. Hussain, "A review of shorthand systems: From brachygraphy to microtext and beyond," *Cognitive Computation*, vol. 12, no. 4, pp. 778–792, 2020.

[34] E. Cambria, Q. Liu, S. Decherchi, F. Xing, , and K. Kwok, "SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis," in *LREC*, 2022, pp. 3829–3839.

[35] R. Satapathy, S. Rajesh Pardeshi, and E. Cambria, "Polarity and subjectivity detection with multitask learning and bert embedding," *Future Internet*, vol. 14, no. 7, p. 191, 2022.

[36] H. T. Nguyen, P. H. Duong, and E. Cambria, "Learning short-text semantic similarity with word embeddings and external knowledge sources," *Knowledge-Based Systems*, vol. 182, no. 104842, 2019.

[37] W. Li, L. Zhu, Y. Shi, K. Guo, and E. Cambria, "User reviews: Sentiment analysis using lexicon integrated two-channel CNN–LSTM family models," *Applied Soft Computing*, vol. 94, no. 106435, 2020.

[38] A. Alshahrani, M. Ghaffari, K. Amirizirtol, and X. Liu, "Identifying optimism and pessimism in twitter messages using xlnet and deep consensus," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[39] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.

[40] A. Kumar, T. E. Trueman, and E. Cambria, "Gender-based multi-aspect sentiment detection using multilabel learning," *Information Sciences*, vol. 606, pp. 453–468, 2022.