Unveiling Chronic Stress: A Social Media Perspective Using Machine Learning

[1]Katroth Balakrishna Maruthiram, [2] Pillalamarri Veena
[1]Assistant Professor in CSE, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, India
[2]MCA Student, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, India

Abstract: Stress is a universal experience, manifesting in two forms: acute and chronic. Acute stress arising briefly from events like traffic jams or disagreements, is a natural part of daily life, enhancing our stress response system's vigilance. In contrast, chronic stress results from prolonged exposure to diverse stressors, straining the body's normal functioning and contributing to severe health issues such as heart disease and mental disorders. Recognizing the influence of social media, researchers have explored novel ways to identify chronic stress in users. They compiled data from 971 individuals facing chronic stress, analyzing 54,546 social media posts from Sina microblog between July 5, 2018, and December 1, 2019. The study introduced innovative techniques: A specialized stress-oriented word embedding to enhance sensitivity in detecting stress-related expressions and a sophisticated multi-attention model capturing post interconnections, accurately inferring long-term stress levels and categories. Results were remarkable, with an 80.65% accuracy in detecting category-aware stress levels. 86.49% accuracy in identifying chronic stress levels, and an outstanding 93.07% accuracy in pinpointing chronic stress categories. This pioneering approach not only illuminates the pervasive issue of chronic stress in the digital age but also offers a robust methodology for precise identification, paving the way for targeted interventions and support systems for those grappling with chronic stress.

Keywords – Chronic stress detection, social media, attention mechanism.

1. INTRODUCTION

Everyone experiences stress, which can be classified into two types: acute and chronic. Chronic stress lasts longer than acute stress. Short-term stressors, such as traffic jams, arguments, and performance evaluations, are examples of acute stress, which helps keep our stress response system in check. Chronic stress, however, occurs when we are repeatedly exposed to the same or multiple stressors over time. This prolonged exposure gradually increases our resting heart rate, blood pressure, breathing rate, and muscle tension, making our bodies work harder to function normally. If left unaddressed, chronic stress can lead to serious health issues such as hypertension, heart disease, chronic pain, and depression. In fact, persistent stress is associated with a wide range of health problems. To avoid these complications, it's important to identify and manage chronic stress. This research investigates chronic stress in contexts such as academics, work, family, interpersonal relationships, self-image, peer interactions, and overall life. Social media is utilized in this study to detect chronic stress for two primary reasons. First, social media is an integral part of daily life, offering vast data storage and strong networks that facilitate the exchange of information, thoughts, and emotions. The extensive historical data from social media users can reveal patterns of chronic stress. Second, compared to surveys or physiological monitoring, social media offers a non-invasive, efficient, and wide-reaching method. Specifically, the language used in social media posts can indicate levels of stress. This study employs a classification-based approach to identify chronic stress related to academics, work, family, personal relationships, self-perception, peer interactions, and overall life by analyzing a series of linguistic posts over time.

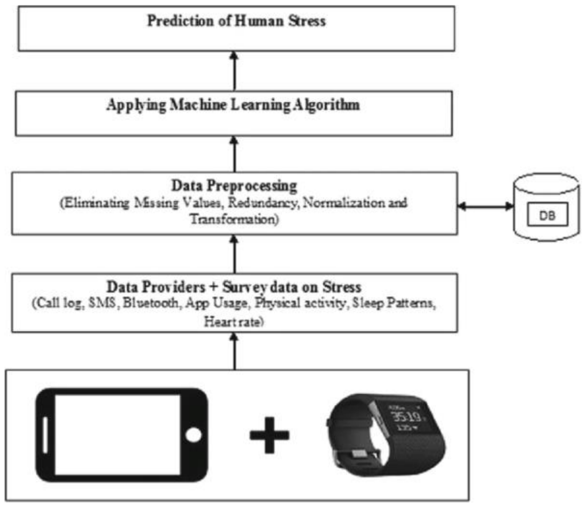

Fig.1: Example figure

The vast number of historical recordings provided by social media users allows us to detect their long-term chronic stress. Second, as compared to systems that use questionnaires or other physical sensors to detect physiological signs, social media are non-contact, labor-inefficient, and have a large reach. We may learn certain stress categories, in particular, from users' linguistic social media postings. In order to make the greatest use of social media resources, we conducted a category-aware chronic stress detection task in this research. That is, we propose to identify user's chronic stress levels in stress categories of study/work, family, intimate connection, self-cognition, peer relation, and life over the period using a list of user's daily postings, each of which comprises a series of linguistic words

OBJECTIVE OF PROJECT

The purpose of this research is to identify chronic stress in the areas of study, job, family, interpersonal relationships, self-cognition, peer relationships and life.

2. LITERATURE REVIEW

Effects of stress on heart rate complexity-A comparison between short-term and chronic stress:

This research compared the effects of chronic and short-term stress on heart rate variability (HRV) in 50 healthy people, comparing time, frequency, and phase domain (complexity) assessments. Chronic stress was measured using the hassles frequency subscale of the combined hassles and uplifts scale (CHUS). A speech task was used to test short-term stressor reaction. Surface electrocardiograms(ECG) were used to evaluate HRV. Because the rate of breathing decreased throughout the speech task (p.001), this research investigated the impact of respiration rate alterations on the effects of interest. Short-term stress lowered HR D2 (measured using the pointwise correlation dimension PD2) (p.001), but raised HR mean (p.001), standard deviation of R-R (SDRR) intervals (p.001), low (LF) (p.001), and high frequency band power (HF) (p =.009). Short-term stress had no effect on respiratory sinus arrhythmia (RSA) or the LF/HF ratio. HR D2 was linked with chronic stress (r =.35, p =.019) when the partial correlation was adjusted for respiration rate. Chronic and short-term stress had different impacts on many HRV metrics. HR D2 dropped under both stress circumstances, indicating that the cardiac pacemaker was less functioning. The findings support the usefulness of complexity measures in contemporary HRV stress studies.

Chronic stress, acute stress, and depressive symptoms

Although life events remain the primary focus of stress research, current results show that chronic stress should be prioritized. Data from a large community survey of married men (n = 819) and women (n = 936) are used to evaluate this topic. In all but one life area, the results demonstrate that chronic stress is more significantly connected to depressed symptoms than acute stress. Chronic and acute stress interaction patterns are mostly related with lower levels of depression than indicated by a main effects model. This trend implies that persistent stress may mitigate the emotional impacts of acute stress. Although the mechanisms behind this impact are unknown, it is hypothesized that anticipation and reassessment lower the stressfulness of an event by making its meaning more benign. The implications for future study on the effects of chronic and acute stress are highlighted.

A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health

Affective computing is becoming increasingly popular as people become more aware of the link between emotional moods and physical health. Emotional computing detects a person's affective state using both hardware and software technologies. It is a dynamic study topic with significant advancements in technology focused toward affective state analysis. Dr. Rosalind Picard of the Massachusetts Institute of Technology (MIT) is credited with coining the term in her 1995 work on affective computing [1]. Since then, it has evolved into a contemporary discipline of computer science for human-computer interactions [2], [3]. This branch of computer science is divided into two sections: 1) detection and identification of emotional information and 2) simulation of emotion in computational systems. The present study focuses on the detection and identification of emotions as affective states.

Towards a microblog platform for sensing and easing adolescent psychological pressures

Adolescent mental health cannot be overlooked, and psychological strain is one of today's most pressing issues for teens. Because of its unique equality, freedom, fragmentation, and individuality characteristics, micro-blogging, as the most important information exchange and broadcast tool in today's society, is becoming an important channel for teenagers' information acquisition, inter-interaction, self-expression, and emotion release. This poster depicts a micro-blog platform that will (1) detect psychological stressors in teens' tweets and (2) let them relieve their tension through micro-blog. A strategy for identifying psychological stressors in teens' tweets is detailed in detail. Our early experimental findings on real-world data confirm the approach's validity. At the conclusion of the poster, we also explore strategies to help teens vent their stress via micro-blogging.

Detecting adolescent psychological pressures from micro-blog

Adolescents are subjected to a variety of psychological stresses as a result of their studies, communication, attachment, and self-recognition. If these psychological stresses are not appropriately addressed, they may manifest as mental issues, which may have catastrophic implications. Due to a lack of timeliness and variety, traditional face-to-face psychological diagnosis and therapy cannot satisfy the need of totally easing teens' stress. With micro-blogging becoming a popular media channel for teenagers' information acquisition, interaction, self-expression, and emotion release, we envision a micro-blogging platform that can detect psychological pressures in teenagers' tweets and help teenagers release their stress through micro-blogging. We study a variety of variables that may indicate teens' stress levels from their tweets and then evaluate five classifiers for pressure identification (Naive Bayes, Support Vector Machines, Artificial Neural Network, Random Forest, and Gaussian Process Classifier). We also show how to aggregate single-tweet detection findings in time series to see how teens' stress fluctuates over time. The Gaussian Process Classifier provides the best detection accuracy because to its resilience in the face of a high degree of uncertainty that may be met with previously unreported training data on tweets, according to experimental findings. Among the elements, the tweet's emotional degree, which combines negative emotional terms, emojis, exclamation and question marks, is crucial in detecting psychological strain.

3. METHODOLOGY

- i) Proposed Work: In our proposed system, we have implemented a Voting Classifier and a CNN + LSTM model. The Voting Classifier performed particularly well, achieving accuracies of 94.95% and 93%.
- The model built with the Voting Classifier is used to predict outcomes based on features selected using specific algorithms.
- For the front end of this project, we will use the Flask framework as an extension. Additionally, user authentication will be provided using an SQLite3 database.

Advantages:

1. The proposed multi-attention model with stress oriented word embedding can identify category-aware stress levels with high accuracy, chronic stress levels with high accuracy, and chronic stress categories with high accuracy.
2. A multi-attention model was then developed to collect inter-relationships from a series of user postings, with the goal of estimating users' long-term category-specific stress levels.

ii) System Architecture:

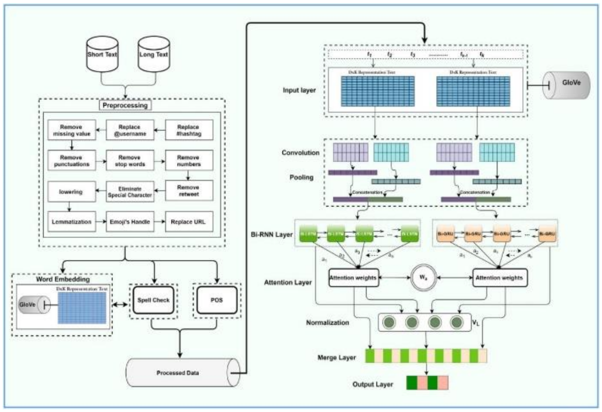

Fig.2: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Data exploration: we will put data into the system using this module.
- Processing: we will read data for processing using this module.
- Splitting data into train and test: Using this module, data will be separated into train and test models.
- Attention Model - MAM - BERT - Cross AutoEncoder - CNN + LSTM - SVM - HIS - Logistic Regression for ElasticNet Classification - SVM - Voting Classifier (Extension) - RNN - LSTM - Guassian Process TeenSensor. Calculated algorithm accuracy.
- User signup and login: Using this module will result in registration and login.
- User input: Using this module will result in predicted input.
- Prediction: final predicted shown

4. IMPLEMENTATION

ALGORITHMS:

Attention Model – MAM:

Attention models, also known as attention mechanisms, are neural network input processing strategies that enable the network to concentrate on certain parts of a complicated input one at a time until the whole dataset is classified. The idea is to divide complex activities into smaller regions of focus that are processed sequentially. Similar to how the human mind handles a new issue by breaking it down into smaller tasks and tackling them one at a time.

BERT:

BERT (Bidirectional Encoder Representations from Transformers) is a deep learning model in which every output element is linked to every input element and the weightings between them are dynamically determined depending on their relationship.

BERT Model

```
def create_model(bert_model):
    input_ids = tf.keras.Input(shape=(60,),dtype='int32')
    attention_masks = tf.keras.Input(shape=(60,),dtype='int32')

    output = bert_model([input_ids,attention_masks])
    output = output[1]

    output = tf.keras.layers.Dense(32,activation='relu')(output)
    output = tf.keras.layers.Dropout(0.2)(output)

    output = tf.keras.layers.Dense(1,activation='sigmoid')(output)
    model = tf.keras.models.Model(inputs = [input_ids,attention_masks],outputs = output)
    model.compile(Adam(lr=6e-6), loss='binary_crossentropy', metrics=['accuracy'])
    return model
```

Cross AutoEncoder:

An autoencoder is a form of artificial neural network that uses machine learning to learn efficient codings of unlabeled input (unsupervised learning). [1] By trying to recreate the input from the encoding, the encoding is checked and enhanced. By training the network to disregard inconsequential input ("noise"), the autoencoder learns a representation (encoding) for a collection of data, generally for dimensionality reduction.

Cross AutoEncoder

```
from keras.layers import Input, Dense, Bidirectional, GRU, Embedding, Dropout, LSTM
from keras.layers import concatenate, SpatialDropout1D, GlobalAveragePooling1D, GlobalMaxPooling1D
from keras.models import Model
from keras import regularizers

epochs=20

# Input shape
inp = Input(shape=(100,))

encoder = Embedding(5000, 100)(inp)
encoder = Bidirectional(LSTM(75, return_sequences=True))(encoder)
encoder = Bidirectional(LSTM(25, return_sequences=True,
                        activity_regularizer=regularizers.l1(10e-5)))(encoder)

decoder = Bidirectional(LSTM(75, return_sequences=True))(encoder)
decoder = GlobalMaxPooling1D()(decoder)
decoder = Dense(50, activation='relu')(decoder)
decoder = Dense(num_classes)(decoder)

model = Model(inputs=inp, outputs=decoder)
model.compile(loss='mean_squared_error', optimizer='adam', metrics=['accuracy'])
```

Logistic Regression for ElasticNet Classification:

The logistic model (or logit model) is a statistical model that represents the likelihood of an event occurring by making the event's log-odds a linear combination of one or more independent variables. Logistic regression[1] (or logit regression) in regression analysis is used to estimate the parameters of a logistic model (the coefficients in the linear combination). In binary logistic regression, there is a single binary dependent variable, coded by an indicator variable, with two values labelled "0" and "1," and the independent variables might be binary variables (two classes, coded by an indicator variable) or continuous variables (any real value).

SVM:

Support Vector Machine, or SVM, is a prominent Supervised Learning technique that is used for both classification and regression issues. However, it is mostly utilized in Machine Learning for Classification difficulties. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary.

Support Vector Machine - HIS

```
from sklearn.svm import SVC
svm = SVC(gamma='auto')
svm.fit(X_train, y_train)
y_pred = svm.predict(X_test)
print(confusion_matrix(y_test,y_pred))
print(classification_report(y_test,y_pred))
print(accuracy_score(y_test, y_pred))
svm = accuracy_score(y_test, y_pred)
```

Voting Classifier (Extension):

A voting classifier is a machine learning estimator that trains many base models or estimators and predicts by aggregating their results. Aggregating criteria may be coupled voting decisions for each estimator output.

RNN - LSTM :

Unlike traditional feedforward neural networks, LSTM has feedback connections. A recurrent neural network (RNN) of this kind may analyze not just single data points (such as photos), but also complete data sequences (such as speech or video).

CNN - LSTM :

CNN + LSTM

```
from keras.models import Sequential
from keras.layers import Dense,Conv1D,MaxPooling1D
from keras.layers import LSTM,Dropout
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence
from keras.callbacks import ModelCheckpoint
```

```
model = Sequential()
model.add(Embedding(5000, 100, input_length=100))
model.add(Conv1D(filters=32, kernel_size=3, padding='same', activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(LSTM(100))
model.add(Dense(num_classes, activation='sigmoid'))
model.compile(loss=tf.keras.losses.SparseCategoricalCrossentropy(), optimizer='adam', metrics=['accuracy'])
print(model.summary())
filepath="weights_best_cnn.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='val_acc', verbose=1, save_best_only=True, mode='max',save_weights_only=True)
callbacks_list = [checkpoint]
```

5. EXPERIMENTAL RESULTS

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\textit{True Positive}}{\textit{True Positive} + \textit{False Positive}}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

Accuracy: Accuracy is the proportion of correct predictions in a classification task, measuring the overall correctness of a model's predictions.

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

F1 Score: The F1 Score is the harmonic mean of precision and recall, offering a balanced measure that considers both false positives and false negatives, making it suitable for imbalanced datasets.

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$

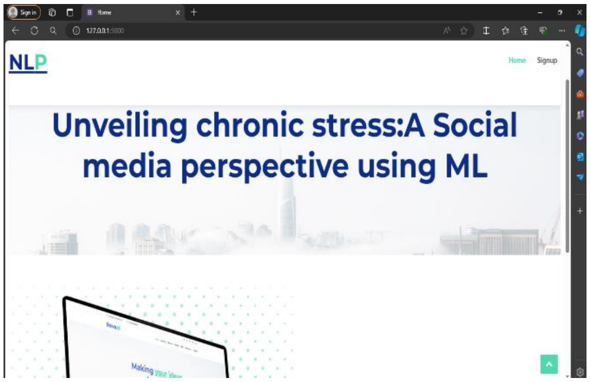

Fig.3: Home screen

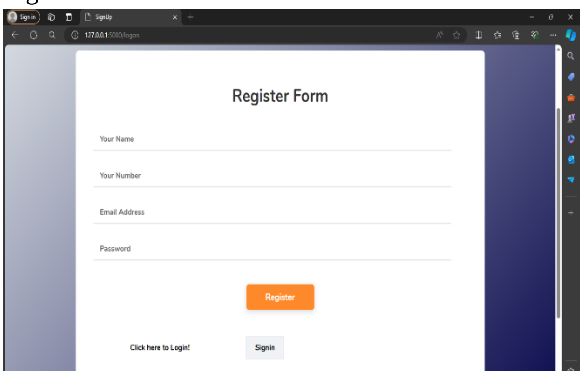

Fig.4: User registration

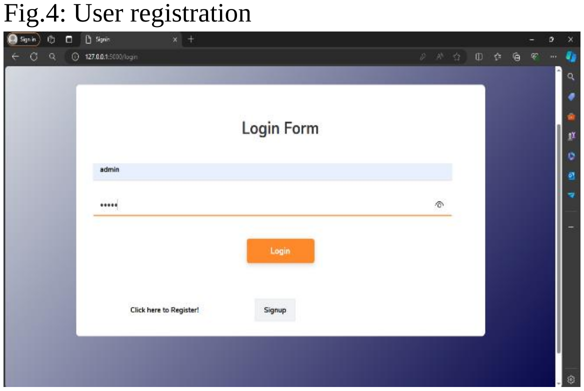

Fig.5: user login

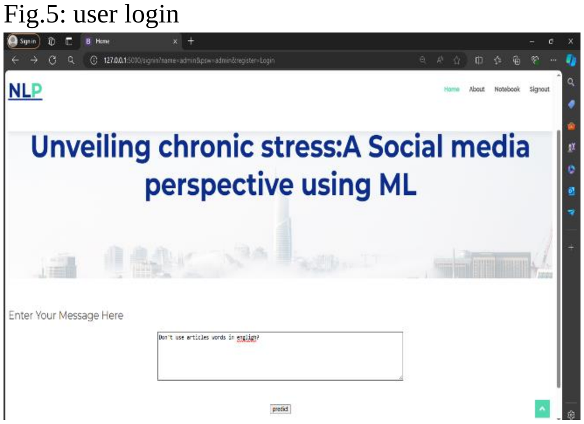

Fig.6: Main screen

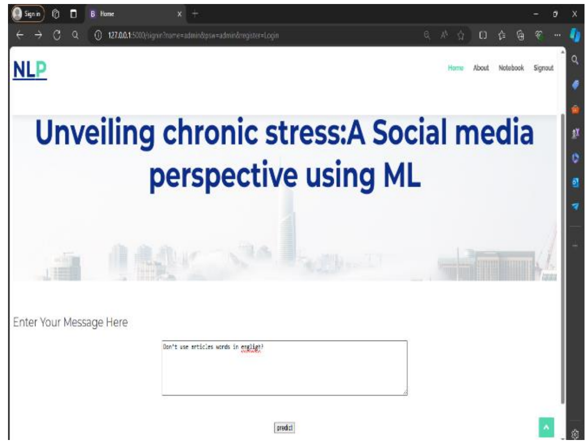

Fig.7: User input

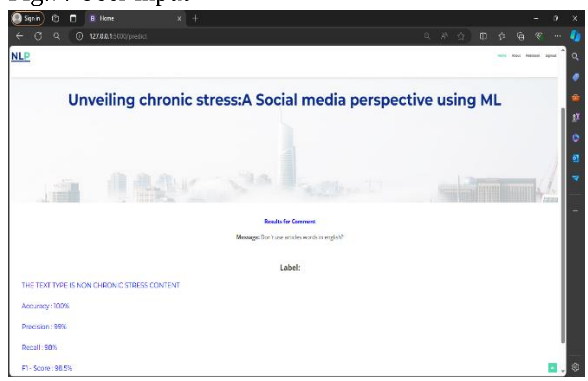

Fig.8: Prediction result

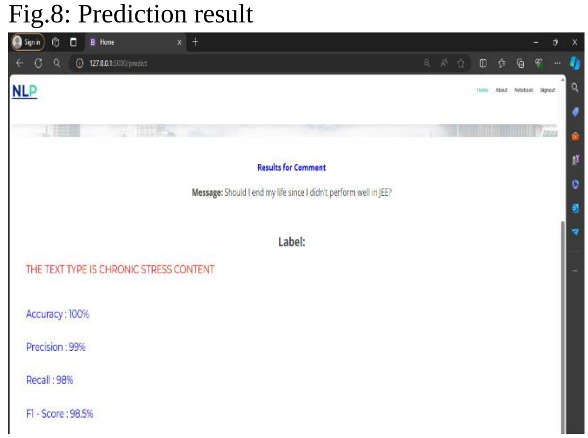

Fig.9: Negative prediction result

6. CONCLUSION

For Unveiling chronic stress: A social media perspective, we built a stress-oriented word embedding to strengthen the sensibility of stress-related expressions for linguistic post analysis.

A multi-attention model was further designed to capture inter-relationships from a sequence of users' posts, aiming to infer users' long-term category-specific stress levels.

Our performance study on the 971 microblog users (392 strongly stressed and 579 weakly stressed) shows that the proposed Voting Classifier, CNN+LSTM and Multi attention model equipped with the stress-oriented word embedding can achieve 94.95% and 93% of accuracy in category-aware chronic stress detection.

FUTURE SCOPE

We want to evaluate and infer user kinds from social media in the future for customized fine-grained chronic stress detection. We investigate this predicted result's accuracy and error rate. For subsequent analysis based on prediction accuracy, we employ the model with the highest score.

REFERENCE

[1] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale, “Effects of stress on heart rate complexity-A comparison between short-term and chronic stress,” Biol. Psychol., vol. 80, no. 3, pp. 325–332, 2009.

[2] K. A. McGonagle and R. C. Kessler, “Chronic stress, acute stress, and depressive symptoms,” Amer. J. Community Psychol., vol. 18, no. 5, pp. 681–706, 1990.

[3] R. R. Baheti and S. Kinariwala, “A survey: Stress detection techniques,” Int. J. Appl. Evol. Comput., vol. 11, no. 1, pp. 28–33, 2020.

[4] S. Greene, H. Thapliyal, and A. Caban-Holt, “A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health,” IEEE Consum. Electron. Mag., vol. 5, no. 4, pp. 44–56, Oct. 2016.

[5] Y. Xue, Q. Li, L. Feng, G. D. Clifford, and D. A. Clifton, “Towards a microblog platform for sensing and easing adolescent psychological pressures,” in Proc. ACM Conf. Pervasive Ubiquitous Comput. Adjunct Pub., 2013, pp. 215–218.

[6] Y. Xue, Q. Li, L. Jin, L. Feng, D. A. Clifton, and G. D. Clifford, “Detecting adolescent psychological pressures from micro-blog,” in Proc. Int. Conf. Health Inf. Sci. Cham: Springer, 2014, pp. 83–94.

[7] L. Zhao, Q. Li, Y. Xue, J. Jia, and L. Feng, “A systematic exploration of the micro-blog feature space for teens stress detection,” Health Inf. Sci. Syst., vol. 4, no. 1, pp. 1–12, 2016.

[8] N. Zhao, D. Jiao, S. Bai, and T. Zhu, “Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services,” PLoS One, vol. 11, no. 6, 2016, Art no. e0157947.

[9] H. Lin et al., “User-level psychological stress detection from social media using deep neural network,” in Proc. 22nd ACM Int. Conf. Multimedia, 2014, pp. 507–516.

[10] M. Thelwall, “Tensistrength: Stress and relaxation magnitude detection for social media texts,” Inf. Process. Manage., vol. 53, no. 1, pp. 106–121, 2017