Received 16 February 2024, accepted 27 February 2024, date of publication 4 March 2024, date of current version 14 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3373558

RESEARCH ARTICLE

Attention and Meta-Heuristic Based General Self-Efficacy Prediction Model From Multimodal Social Media Dataset

MD. SADDAM HOSSAIN MUKTA[1], **JUBAER AHMAD**[2], **AKIB ZAMAN**[3], **AND SALEKUL ISLAM**[2], **(Senior Member, IEEE)**

[1]LUT School of Engineering Sciences, Lappeenranta-Lahti University of Technology, 53850 Lappeenranta, Finland
[2]Department of Computer Science and Engineering, United International University (UIU), Dhaka 1212, Bangladesh
[3]Computer Science and Artificial Intelligence Laboratory (CSAIL), Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA

Corresponding author: Md. Saddam Hossain Mukta (Saddam.Mukta@lut.fi)

This work was supported by United International University, Bangladesh, under Grant UIU-IAR-01-2022-SE-06.

ABSTRACT General Self-Efficacy (GSE) is a vital attribute of human psychology that describes one's belief about his own ability to succeed in specific situations. GSE is composed of cognitive, social, and behavioral skills of an individual. In this research, we first develop a GSE classification model by using Facebook content (i.e., profile photos and statuses). We collect data from a total of 435 Facebook users in an ethical data collection manner. Two hybrid machine learning methods are applied based on distinct feature extraction approaches: tool-based and deep learning-based. In our tool-based approach, we employ Linguistic Inquiry and Word Count (LIWC) and Bidirectional Encoder Representations from Transformers (BERT) for text and Mediapipe and DeepFace for image feature extraction. We apply Particle Swarm Optimization (PSO) for feature selection, resulting in a robust tabular dataset with high predictive performance for GSE scores. In the deep learning-based approach, we apply BERT and 1-dimensional convolutional neural network (1D-CNN) for text feature extraction, while UNet++ handles image segmentation, and VGG16 and ResNet-152 contribute image features, fused via Canonical Correlation Analysis (CCA). We also integrate a co-attention model for image and text features. Traditional machine learning models, including Random Forest (RF), Xgboost, AdaBoost, and Stacking, are then trained on the feature set to predict GSE scores. This comprehensive model showcases a multifaceted approach to GSE prediction, combining tool-based and deep learning methodologies for enhanced accuracy and insights. Then, we develop a GSE prediction model by using the mentioned tool-based (i.e., LIWC, BERT, Mediapipe, and DeepFace) and deep learning-based feature extraction methods from both image and text datasets. The tool-based model achieves remarkable accuracy percentages of 85.80% (text), 91.06% (image), and an outstanding 93.25% for the hybrid model. The deep learning-based model exhibits competitive results, with accuracies of 64.80% (text), 73.06% (image), and 81.87% for the hybrid model.

INDEX TERMS Self-efficacy, deep learning, convolutional neural network (CNN), artificial intelligence (AI), Facebook, multimodal dataset, PSO, machine learning, co-attention, segmentation, classifiers.

I. INTRODUCTION

In recent times, social media have become an integral part of the lives of millions of users. Users express their opinions, ideas, and preferences through these social media sites with their friends, family, colleagues, and acquaintances. Researchers identify different psychological attributes such as personality [1], values [2], and emotion [3] by analyzing the contents posted in social media. Among different psychological attributes, *self-efficacy* is an important attribute,

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Huang.

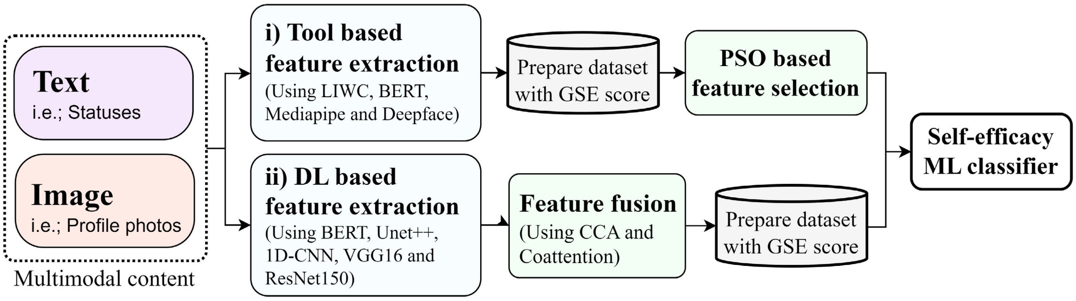

Text
i.e.; Statuses
Image
i.e.; Profile photos
Multimodal content
i) Tool based feature extraction
(Using LIWC, BERT, Mediapipe and Deepface)
ii) DL based feature extraction
(Using BERT, Unet++, 1D-CNN, VGG16 and ResNet150)
Prepare dataset with GSE score
PSO based feature selection
Feature fusion
(Using CCA and Coattention)
Prepare dataset with GSE score
Self-efficacy ML classifier

FIGURE 1. High-level architecture of the self-efficacy prediction model from multimodal social media content.

which describes one's belief in his ability to accomplish a task [4]. In this paper, we predict users' *general self-efficacy (GSE)* from their multimodal content (i.e., statuses and profile photos) posted on Facebook by using machine learning models.

Self-efficacy [4] is a belief of an individual, which is composed of cognitive, social, and behavioral skills to be successful in specific situations. A person with high score in *self-efficacy* has strong judgmental process to execute a course of actions to accomplish a task. For example, a student with high *self-efficacy* may not be competent in a course, but she may learn the course well and obtain a good score in the final examination. A fresh graduate with high *self-efficacy* may attend a complex problem solving group and can contribute to the group effectively.

In recent years, several studies have been conducted to find many interesting insights from the social media, e.g., Facebook contents [1], [5]. A few studies have been conducted to predict users' *self-efficacy* from the social media. Wang et al. [6] conduct a study to investigate the role of general and specific computer *self-efficacy* for measuring the performance of Facebook usage from different features such as posting a status, poking a person, and uploading a photo. Depending on the scores of *self-efficacy* in using these Facebook features, the authors predict whether these users continue using Facebook in the future. Compeau and Higgins [7] also discuss the role of *self-efficacy* to use computers competently. The majority of these studies have been conducted to predict *self-efficacy* in the perspective of socio-psychological aspects.

Schwartz et al. [8] demonstrate that people express their intimate information in Facebook more comfortably than in their real-life. Back et al. [5] show that people reveal their actual personality and behavior in Facebook. Eftekhar et al. [9] show that users represent themselves by their profile photos and writings. Motivated by these studies [5], [8], in this paper, we focus to identify users' *self-efficacy* from their Facebook content. Determining *self-efficacy* of a user has a number of real life applications. For example, employers and decision makers can identify the appropriate person for their organization, people can elect the potential leader for their society, and stakeholders can also find skilled people for their business. To the best of our knowledge, our study is the first to predict users' *self-efficacy* by analyzing their textual and visual contents derived from their Facebook usage which produces better comprehension about a user.

A high-level architecture of our self-efficacy prediction model from multimodal content (i.e., text and image) of social media is shown in Figure 1. In developing the prediction model, we first collect statuses of a total of 435 active Facebook users. We conduct a gold-standard survey of 10-items test among these 435 users as ground truth data on the GSE scores. To predict the GSE score from the available dataset, we design two key hybrid machine learning methods depending on different feature extraction approaches: *tool-based feature extraction* and *deep learning-based feature extraction*. In the tool-based approach, we utilize Linguistic Inquiry and Word Count *(LIWC)* [10] and Bidirectional Encoder Representations from Transformers (*BERT*) [11] for text feature extraction, and *Mediapipe* and *DeepFace* for image feature extraction. We merge these features and select relevant features by using a Particle Swarm Optimization (PSO)-based metaheuristic algorithm [12]. This approach results in the creation of a tabular dataset that demonstrate strong predictive performance in determining users' GSE scores.

On the other hand, in the deep learning-based approach, we employ BERT for word embedding and 1-dimensional convolutional neural network (1D-CNN) for text feature extraction. For image feature extraction, we use UNet++ [13] for image segmentation and extracted features from VGG16 and ResNet-152, which are then fused using Canonical Correlation Analysis (CCA) [14]. Finally, a co-attention model is used to fuse the image and text features for predicting users' GSE scores. After that, we train the feature set and predict the GSE score using traditional machine learning models (i.e., Random Forest (RF), Xgboost, AdaBoost, and Stacking [15]).

In summary, the salient contributions of this study can be summarized as follows:

- We are the first (to the best of our knowledge) to predict users' *self-efficacy* from their multimodal content, i.e., statuses and profile photos in Facebook, by using machine learning techniques.

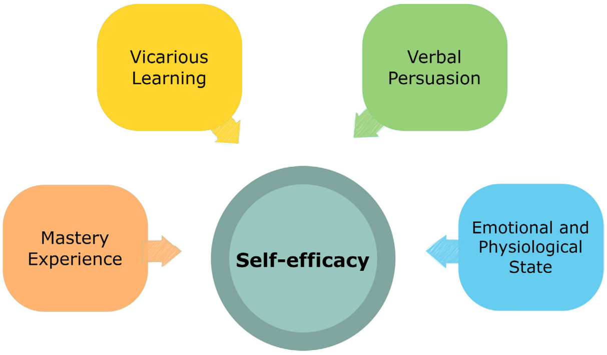


FIGURE 2. Self-efficacy components from human behaviour.

- We present two advanced feature extraction methods: i) tool-based and ii) deep learning-based, to predict users' self-efficacy level by using traditional machine learning based classifiers.
- We implement a self-attention based co-attention method that combines the features of images and textual data.
- We develop a self-efficacy dataset from social media by following Bandura's theory [16].

The rest of the paper is structured as follows: section II presents previous studies related to self-efficacy, section III outlines the data collection and dataset preparation processes, section IV ans section V focuses on the model development procedure, section VI, presents a comparative analysis of our self-efficacy identification using two distinct methods. The findings are then elaborated in section VII and the paper is finally concluded.

II. BACKGROUND AND LITERATURE REVIEW

In this section, we first present the background information and other essential components of self-efficacy. We divide the related studies into two different areas: the prediction of human psychological attributes from social media and self-efficacy in terms of socio-psychological perspective.

A. SELF-EFFICACY

The term *self-efficacy* was first introduced by psychologist Albert Bandura [16] in 1977. Self-efficacy is the faith that a person has in her capacity to carry out tasks successfully and produce desired results. It is an idea in psychology that has an effect on inspiration, setting goals, resiliency, and general performance. Self-reports, behavioral evaluations, and performance assessments are used to determine a person's level of confidence in evaluating self-efficacy parameters.

Bandura's theory [17] of self-efficacy has provided us with a widely recognized psychological framework which consists of four essential components. Figure 2 presents those four components which influence the nature of self-efficacy of an individual.

1) MASTERY EXPERIENCES

Mastery experiences refer to past experiences of successfully performing tasks or achieving desired outcomes. These experiences have a direct impact on an individual's self-efficacy. When individuals have positive mastery experiences and they have successfully accomplished a task, it enhances their level of self-efficacy. Successes and accomplishments provide concrete evidence that individuals can effectively perform specific tasks and achieve desired outcomes. On the other hand, if individuals have negative mastery experiences, such as repeated failures or difficulties in performing tasks, it can diminish their self-efficacy. Unsuccessful attempts and setbacks can create doubts and lower confidence in one's abilities.

2) VICARIOUS LEARNING

Self-efficacy derived from vicarious experiences involves observing others' behavior in specific tasks and assessing one's own capability to replicate those behaviors. By observing and comparing oneself to individuals with similar or slightly higher abilities, individuals can enhance their self-efficacy through modeling and social comparisons. Vicarious experiences can also be obtained from media models, such as videos, that provide opportunities for observational learning.

3) VERBAL PERSUASION

Verbal feedback plays a significant role in shaping self-efficacy as it provides individuals with information about their task performance. Positive feedback, in particular, contributes to the belief that one is capable of performing well. The impact of feedback is enhanced when it is delivered by a credible source, someone whose opinion is respected and trusted by the recipient. Detecting self-efficacy involves assessing the impact of positive or negative feedback, encouragement, and supportive communication on individuals' beliefs about their capabilities.

4) EMOTIONAL AND PHYSIOLOGICAL STATES

In Bandura's theory [16], positive emotional states and low levels of anxiety are associated with higher level of self-efficacy. When individuals experience positive emotions such as confidence, enthusiasm, and optimism, they are more likely to believe in their own capabilities to succeed in a given task. Conversely, high levels of anxiety, fear, or stress can diminish self-efficacy as they may lead to self-doubt and a lack of confidence in one's abilities. Physiological states, such as changes in heart rate, sweating, or other bodily responses, can also influence self-efficacy.

B. PREDICTION OF PSYCHOLOGICAL ATTRIBUTES FROM SOCIAL MEDIA CONTENTS

Several studies [1], [3], [18] have shown a strong connection between a person's cognitive attributes and his/her interactions in social networking sites (SNS). Commonly, users express their views, self-beliefs and characteristics through both writing and non-verbal (i.e., photos) way of communication in social media. They are vocal on these sites from socio-cultural to political [3] issues. Through these digital records, researchers able to predict even users' highly sensitive personal traits such as sexual orientation, ethnicity, religious and political views, intelligence, and so on [1]. Surprisingly, both user-generated and user-supported contents provided in these platforms are even suitable for identifying basic human values [18]. Moreover, people belonging to the same egocentric network often share common personality traits [19]. Several studies also examine the psychological well-being of users based on their Facebook engagements [20] and behavioral patterns derived from their interactions in Twitter [21].

A considerable number of studies [2], [22], [23] advocate the success of predicting personalities and values from social media content. Khan et al. [22] predict users' preferred movie genre based on their personality traits and values derived from their tweets. Golbeck et al. [23] perform a regression analysis between publicly available users' Facebook content and their Big5 personality traits. Number of friends, egocentric network density, activities and preferences, personal information, and the linguistic features obtained from user's textual information are the independent variables for predicting personality traits. The word usage patterns of a user can also be influenced by his/her five higher-level values [2]. For this experiment, the authors choose the participants from Reddit, which is one of the most popular social media platforms. They compute the linguistic features from the posts of the users by using the LIWC tool and also conduct an assessment among them using Portrait Values Questionnaire (PVQ) for their value scores. They authors perform both regression and classification analysis on the collected data. The correlation between a user's general reading interest and the PVQ values can be seen in a different study [24].

Again, photos shared in SNS can also be a prospective dataset for identifying different human traits [25], [26], [27]. Celli [25] introduces a classification technique to recognize users' personality traits and interaction styles through their Facebook profile pictures. The features from the photos were extracted using the Bag-of-Visual-Word (BoVW) method. The author conducted two separate surveys, namely the Big5 personality test and the interpersonal circumplex, to assess the participants' personality traits and interaction styles. A similar study [26] conducted on Twitter users shows that people belonging to different personality traits likely post photos conveying different messages. Agreeable and conscientious users express more positive emotions in their photos, whereas, users high in openness prefer to post more aesthetic photos. Visual features in photos can also vary in respect to different personalities [27]. Agreeable and extroverted people tend to post warm-colored photos and photos with many faces on Facebook. On the other side, neurotic individuals mostly prefer to post indoor photos.

C. SELF-EFFICACY IN PSYCHOLOGY

Self-efficacy is a person's particular set of beliefs in her capability to execute courses of actions required for a specific performance in a given situation [16], [32]. Several psychological studies [4], [16], [33], [34] assume that self-efficacy changes depending on the situation and surroundings of an individual. Moreover, the sense of self-efficacy in humans can be developed by the influences of mastery experiences, vicarious experiences, verbal persuasion and physiological and affective states [32], [34]. According to Bandura [4], individuals with higher levels of self-efficacy tend to demonstrate greater performance achievements and exhibit reduced emotional reactions. Thus, people with a strong sense of efficacy with little effort can overcome failure and formulate constructive approaches to handle new or difficult situations.

Inspired by the mechanism of self-efficacy several studies have been executed. Compeau et al. [7] construct Computer Self-Efficacy (CSE) in accordance with Bandura's concept. Marakas et el. [35] propose a framework for a generalized form of CSE (GCSE). Furthermore, Marakas et el. [36] present a study with detailed comparisons between the various forms of CSE and GCSE measures through hypothesis testing using statistical validation methods. Additionally, Schwarzer et el. [37] develop a 10-item scale for GSE that appears to be a universal measurement yielding meaningful relations with other psychological attributes [38].

Again, various studies establish a relationship between different forms of self-efficacy and varied aspects of life. McQuiggan et al. [39] examine the association between self-efficacy and a student's physiological state in a learning environment. They propose a multi-class classification technique utilizing students' demographic, problem-solving, and physiological data (heart rate and skin conductance) for predicting different levels of self-efficacy using conventional machine learning models with a moderate performance of 72% (Naive Bayes) and 83% (Decision Tree). In contrast, Wang et al. [6] conduct a study to investigate the role of GCSE and specific CSE for predicting whether a user will continue using Facebook in the future based on the functionality of different features available at Facebook. The authors analyze several hypotheses using statistical methods. As stated earlier, social media can be an informative source for an individual's personality profiling. Similarly, Ouirdi et. al [40] demonstrate a hypothetical model to predict job seekers' self-disclosure on social media (i.e., career-oriented) based on hyperpersonal computer-mediated communication theory, self-efficacy theory, and social exchange theory in contrast to age, gender, education level and work experience, respectively.

Conversely, we have explored several studies that approach the prediction of self-efficacy through diverse machine learning techniques. The paper [31] conducts a case study for predicting a comparative self-efficacy between students and teachers by using decision tree (DT), k-nearest neighbor (kNN), Naïve Bayes (NB), and support vector machine

TABLE 1. Comparison of the different state-of-the-art paper with our work.

Paper	GSE prediction from social media text	GSE prediction from social media Image	Co-attention between text and images	Used Multi-modal dataset	Used Meta-heuristic approach	Tool-based feature extraction	DL-based feature extraction	Used ML model	Used Tree-based model
[28]	×	×	×	×	×	×	×	✓	×
[29]	×	×	×	×	×	×	×	✓	×
[30]	×	×	×	×	×	×	×	✓	✓
[31]	×	×	×	×	×	×	×	✓	×
Our	✓	✓	✓	✓	✓	✓	✓	✓	×

(SVM). The target of the study is to build a prediction model for measuring the performance of students in an introductory programming course which reflect their self-efficacy. This study not only provides valuable insights into predicting at-risk students in programming languages but also contributes to the comparative analysis of applied models through diverse evaluation metrics. Karataş et al. [29] investigates the connection between teacher candidates' academic self-efficacy, self-directed learning, and future time perspective using machine learning classification algorithms, emphasizing the significance of self-directed learning and future time perspective in predicting academic self-efficacy. Tan et al. [30] contributes by using machine learning approaches to model self-efficacy in a vast dataset from the Programme for International Student Assessment (PISA) 2018, showcasing the utility of tree-based ensemble learning models in identifying low self-efficacious learners on a large scale. Finally, Yildiz et al. [28] finds the predictive power of academic engagement, burnout, and proactive strategies on academic self-efficacy among pre-service teachers, employing both linear regression and artificial neural networks (ANNs). Overall, these studies apply different machine learning techniques in predicting self-efficacy across different educational contexts and student populations.

Table 1 provides a comparative analysis between our work and different state-of-the-art papers in the field of self-efficacy prediction. Our work stands out by incorporating multi-modal datasets from social media interactions, utilizing meta-heuristic approaches, and employing both tool-based and DL-based feature extraction methods. Unlike the previous studies, our approach integrates co-attention between text and images, emphasizing a holistic understanding. While the majority of the compared papers employ tree-based models, our work utilizes a broader range of machine learning models. The distinctive features of our approach contribute to a comprehensive and advanced GSE prediction framework.

In the light of the above discussion, it is observed that self-efficacy was not predicted in the domain of machine learning by analyzing social media content. In addition to this, a user's integrated contents (i.e., both textual and visual) may provide comprehensive profile of that user. Our study takes this opportunity and predicts users' self-efficacy by using meta-heuristic and deep learning based methods that use multimodal content from social media interactions.

III. DATA COLLECTION AND PRE-PROCESSING

We randomly invited Facebook users to collect their available Facebook content (both statuses and profile photos). First, we invited 510 users to participate in this study and 435 (male=223 and female=212) users showed interest to share their data as of (August 26, 2022). The participants filled out a consent form willingly and we informed them that their Facebook content will be used for automatic GSE score prediction anonymously. Table 2 presents the statistics of our dataset where the participants are students from engineering, medical science and business backgrounds, and service holders as well. We communicated with the participants through different platforms such as Facebook messenger, email, and telephones to get their data. We collected users' data by using an ethical data collection manner (IREB/2023/005), which was approved by the Institute of Advanced Research (IAR), United International University (UIU), Bangladesh.

TABLE 2. Statistics of our dataset.

Category	Size/Method
Sampling technique	Random Sampling
Communication platforms	Facebook, Email and Telephone
Number of users	435 (M=223, F=212)
Age group	22-40 years
Professions	(Students, bankers, software engineers, IT professionals, physicians, etc.)
Total number of statuses	4,230 (1,45,879 words)
Statuses per participants	10 (avg=10, min=7, max=20 and std=3.6)
Total number of photos	2,493
Photos per participants	(avg=14, min=3, max=25 and std=2.5)
Total size of dataset	871 MB
Class distribution	High: 144, medium: 194 and low: 97

We conduct a gold standard survey by using a 10-item GSE [37] test among the participants. The respondents self-report how they feel when the statements are asked. Each statement describes a situation and a user responds to the items in a 4 point Likert scale. GSE score has a minimum and maximum scores of 10 and 40, respectively. We normalize the scores between 0 to 1. In our study, we apply information entropy to discretize continuous attributes, inspired by the work of Liu et al. [41]. This process involve evaluating entropy, identifying optimal split points, and creating bins or intervals. It enhances data homogeneity and is valuable for analysis and modeling tasks that require discrete input. Then, we label the values based on the entropy based discretization of the GSE scores according to the following range: *high, medium* and *low*. Here, high GSE score is from 0.78 to 1, medium GSE score is from 0.45 to 0.77, and low GSE score is from 0.01 to 0.45. We consider these levels of GSE scores as our ground truth data. After answering all these questions, participants download a compressed file containing their Facebook content.

For cleaning the textual dataset, we convert the emoticons into their semantic forms, e.g., ":)" into `smile`. To replace the emojis, we use the *demoji's*[1] Python implementation package. Later, we remove non-English words from the dataset by using *NLTK's*[2] library. We also eliminate URLs, http links, numbers, and other irrelevant symbols because they do not hold any meaning for our analysis. We do not remove any stop words from the texts because they often express the context better [42]. We also manually check a few posts to ensure the quality. We do not clean our image dataset.

We implement multiple pipelines to train our dataset due to the presence of three different types of data: text, image, and hybrid (text and image). After evaluating all of the implementations, we select two pipelines: *1) tool-based feature extraction method,* and *2) deep learning-based feature extraction method*. In the following sections, we describe the process of designing and building our prediction models based on these two methods.

IV. TOOL-BASED FEATURE EXTRACTION PREDICTION MODEL

In this section, we describe the steps of building our prediction model by extracting features by using tools (i.e., LIWC, BERT, Mediapipe and DeepFace) from both text and image datasets. First, we extract features from the text dataset. Then, we extract features from the image dataset. Next, we integrate both types of features and build a large tabular dataset. Then, we build a classification model from the tabular dataset to predict the GSE score. Figure 3 depicts the general architecture of our tool-based feature extraction prediction model.

A. FEATURE EXTRACTION

This subsection discusses the process of extracting features from text and image data in detail.

1) TEXT DATA

To extract important features from our textual dataset, we conduct two different types of analyses on our cleaned Facebook statuses: i) Psycholinguistic analysis with *LIWC* and ii) Context-based word embedding with *Sentence-Transformers* [11].

a: PSYCHOLINGUISTIC ANALYSIS

First, we analyze users' Facebook statuses by using LIWC2015 [43], which determines approximately 90 different features that are broadly divided into seven major categories. These categories are: summary language variables (analytical thinking, clout), general descriptor (total word count, words per sentence), standard linguistic dimensions (% of words in the text that are pronouns and articles), psychological constructs (affect, cognition), personal concern (work, home), informal language markers (assents, fillers), and punctuation (periods, commas) [10].

b: CONTEXT-BASED WORD EMBEDDING

First, the Facebook statuses are tokenized into words and then converted into feature vectors using the BERT Transformer. Let $s^t = [s^1, s^2,.., s^n]$, where $t \epsilon R^D$ that represents the BERT embedding at position t, n is the length of the status (in terms of words), and D is the dimension of the vocabulary of words. The transformation of s^t into input vectors v^t is achieved through $v^t = W^e s^t$, where W^e is an embedding vector, and v^t is a learned component during training, along with other characteristics. After extracting features from LIWC and BERT, they are merged into one vector, denoted as T.

2) IMAGE DATA

To generate the image features, we use *MediaPipe*[3] and *DeepFace*[4] multi-criteria feature extraction methods. We mainly concentrate on prominent features such as *color, composition, demographic information* and *facial expression*. All the image samples are shaped into 128 × 128 pixels before the feature extraction because we have found several studies [44], [45] that employ similar image shape and achieve excellent results.

a: MEDIAPIPE

For the purpose of identifying facial landmarks and calculating the head posture, Mediapipe Facemesh integrates machine learning and computer vision technologies [46]. Note that facial landmarks are also known as facial key points or facial feature points, which are precise points on a person's face that represent the distinguishing facial characteristics such as the eyes, nose, mouth, and brows. Applications like face identification, emotion analysis, facial expression detection, etc. heavily rely on the detection of these landmarks. The mathematical representation of the facial landmark detection method is a function F that converts an input image x to a collection of 2D face landmarks y. We express y as the outcome of applying function $y = F(x)$, where x is the input image, y represents the set of 2D facial landmarks (values range from 0 to 1), and F is the function executed by the CNN. In our image dataset, we have obtained 468 facial landmark points that correspond to different facial features. Each point represents a unique facial characteristic, and since these points vary across different images, we have used them as features for our analysis. Figure 4 showcases a collection of 468 facial landmark points on a human face. These landmark points are derived from images within our dataset.

[1] https://pypi.org/project/demoji
[2] https://www.nltk.org/howto/wordnet.html
[3] https://mediapipe.dev/
[4] https://viso.ai/computer-vision/deepface/

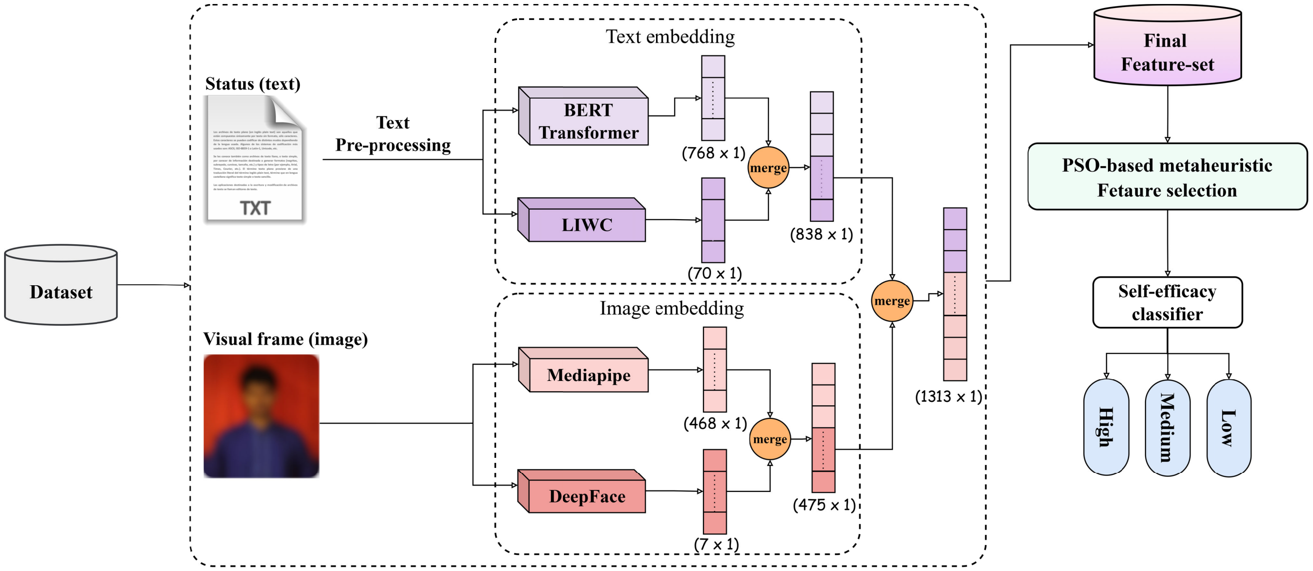


FIGURE 3. Architecture of our GSE prediction model from tool-based feature extraction method.

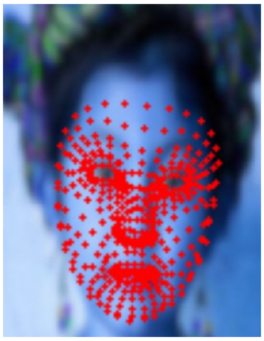


FIGURE 4. Facial landmark points from mediapipe.

b: DEEPFACE

is an advanced deep learning framework created by Facebook's AI Research team [47], specializing in facial analysis tasks. Its primary objectives include facial recognition, facial attribute analysis, facial emotion recognition, and facial landmark detection. The images are processed using a deep neural network to produce a distinctive and concise feature representation known as the deep face descriptor [47]. This feature extraction can be represented as: $y = f(x)$, where x represents the input face image, y represents the deep face descriptor (values range from 0 to 1), and f is the function performed by the CNN. DeepFace library has seven different features based on various emotions: *angry, happy, neutral, sad, surprise, disgust,* and *fear*, which can provide valuable insights to derive human self-efficacy [22], [48]. Table 3 represents a brief description of different features of DeepFace library based on diverse categories of emotion.

TABLE 3. Description of DeepFace features.

Emotion	Description
Angry	Frustration, dissatisfaction, or lack of control. Suggests low self-efficacy.
Happy	Positive emotional state and sense of accomplishment. Indicates high self-efficacy.
Neutral	No specific emotion. Serves as a baseline for comparison.
Sad	Disappointment, helplessness, or lack of control. Suggests low self-efficacy.
Surprise	Sudden and unexpected reaction. Provides insights into response to unexpected events.
Disgust	Aversion or revulsion towards something unpleasant. Suggests low self-efficacy.
Fear	Response to perceived threats or danger. Indicates low self-efficacy.

Table 4 shows the statistical information about the distribution of emotions extracted by DeepFace in our dataset. It includes measures such as the mean, standard deviation, minimum, first and third quartiles, median, and maximum values for each emotion categories. These statistics provide insights of the central tendency, variability, and range of scores for each emotion in the dataset.

The features obtained from both *Mediapipe* and *DeepFace* are combined into a single vector, I. Several studies [49], [50], [51] use multimodal feature fusion techniques that significantly improves results. After using advanced feature extraction methods, we have obtained two feature vectors, one from text data (T) and another from image data (I), which are later merged into a single vector V. Vector V comprises of a total of 1,313 features, including 838 from text data and 475 from image data, (see Table 5). However, as some of these features might not be crucial in determining the GSE score, a feature selection technique is applied to find the relevant feature set.

B. FEATURE SELECTION

In this section, we explore a feature selection method to identify the most significant features from the 1,313 features derived from text and image data during the training

TABLE 4. Statistical scores for different DeepFace features.

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Mean	5.95	0.00	11.03	26.02	13.87	13.87	42.21
STD	15.38	0.02	23.24	39.16	26.90	26.90	40.81
Min	0.001	0.001	0.001	0.001	0.001	0.001	0.001
25%	0.001	0.001	0.001	0.01	0.01	0.01	2.25
50%	0.09	0.00	0.37	0.70	0.77	0.77	23.10
75%	2.48	0.001	8.13	41.90	8.44	8.44	90.30
Max	99.59	0.18	98.43	100.00	99.97	99.97	100.00

TABLE 5. Feature distribution of different tools.

BERT	LIWC	Mediapipe	Deepface
1 ... 768	1 ... 70	1 ... 468	1 ... 7

<——————————— 1313 ———————————>

phase. The PSO-based [52] metaheuristic optimization algorithm is utilized to perform the feature selection process. PSO treats potential solutions as particles in a multidimensional space, adjusting their positions based on personal and neighboring best-known positions. Utilizing PSO for hyperparameter tuning in predicting self-efficacy can be considered advantageous due to its ability to efficiently search through a large hyperparameter space. PSO is a population-based optimization technique inspired by social behavior, enabling it to effectively navigate various hyperparameter configurations. By leveraging PSO, one can potentially enhance the predictive performance of the model by fine-tuning hyperparameters to suit the specific characteristics of the dataset. This approach often leads to better convergence and improved model accuracy compared to traditional grid or random search methods [53], [54].

The fitness function in PSO evaluates a particle's position quality in the search space, guiding their movement toward better solutions. It's represented mathematically as $f(x)$, taking a particle's position x and returning a scalar value indicating the position's fitness. The best fitness model in PSO is the particle with the optimal fitness among all particles. For example, in a 2D scenario aiming to minimize the function $f(x) = x_1^2 + x_2^2$, the particle with the smallest $f(x)$ value among all represents the best fitness model. Here, $x = (x_1, x_2)$ represents the position of a particle in a two-dimensional space. The goal in this example is to minimize the sum of the squares of x_1 and x_2. The fitness function evaluates how "fit" a particular solution (particle's position) is concerning the optimization goal. For a more specific fitness function, we would replace $(x_1^2 + x_2^2)$ with the actual expression that represents the objective or cost function of our optimization problem. This could involve complex mathematical expressions based on the nature of the problem we are trying to solve.

Each particle's location in the search area is updated depending on its present speed as well as the best location discovered itself and its neighbors. The PSO mathematical formulation is described in the following.

Let, $X = [x_1, x_2, \ldots, x_n]$ represents the collection of n particles inside the search area, where x_i represents the i_{th} particle's position and V_i represents its velocity. Let, P_{best} and G_{best} stand for best positions of particle i as well as the swarm as a whole, respectively. Each particle's location and velocity are recalculated as follows:

$$V_{temp1} = WV_i(t) + c_1 \times Rand \times (P_{best} - x_i(t)) \tag{1}$$

$$V_{temp2} = c_2 \times Rand \times (G_{best} - x_i(t)) \tag{2}$$

$$V_i(t+1) = V_{temp1} + V_{temp2} \tag{3}$$

$$x_i(t+1) = x_i(t) + V_i(t+1) \tag{4}$$

where the constants W, c_1, and c_2 control the influence of these factors, and $Rand$ is a randomized number ranging from 0 to 1. The algorithm continues until a termination criterion is met, such as reaching a maximum number of iterations or a satisfactory solution. The velocity and position of each particle in the search space are continually adjusted by the PSO algorithm, taking into account the weight of the previous velocity, attraction towards the particle's personal best position, and attraction towards the global best position found by all particles.

Our dataset for predicting the GSE score consists of complicated feature set, which is a high-dimensional dataset (i.e., a total of 1,313 features). A high dimensional dataset can be handled by PSO [52], since it efficiently explores solution from a wide search space. The technique pinpoints pertinent features or feature clusters that support precise GSE prediction. PSO is renowned for its proficiency in dealing with non-linear and non-convex optimization issues, both of which are typical in behavior (i.e., self-efficacy) prediction tasks.

Table 6 presents the performance of different machine learning models (Random Forest (RF), Xgboost, AdaBoost, and Stacking) on 20 different feature sets derived from the PSO training. As part of our research, we aim to compare and identify the most effective machine learning model among these popular algorithms in handling the given feature sets. By evaluating the fitness scores achieved by each model on different feature sets, we search to establish a baseline for their performance in our specific context. By employing multiple well-established algorithms and testing them on the same feature sets, we aim to gain insights into the strengths and weaknesses of each model for our specific dataset. We find that the results selected through the PSO feature selection method in feature set 14 provide the best fit for the models, resulting in improved accuracy compared to other feature sets. AT the end, we employ the PSO method to choose 219 features, which yields the best performance.

TABLE 6. PSO based fitness score for various machine learning models depending on the different feature sets from tool-based extraction method.

Fetaure set	Random Forest (RF) fitness	Xgboost fitness	AdaBoost fitness	Stacking fitness
1	0.7759	0.8276	0.8448	0.8576
2	0.7586	0.8375	0.8358	0.8776
3	0.7586	0.8046	0.9180	0.8576
4	0.7414	0.8016	0.8239	0.8848
5	0.7759	0.8126	0.8493	0.8448
6	0.7586	0.8245	0.9182	0.8276
7	0.8276	0.8575	0.8504	0.9076
8	0.7414	0.8086	0.8283	0.8948
9	0.7586	0.8060	0.8884	0.8748
10	0.7931	0.7976	0.8244	0.9248
11	0.7586	0.7960	0.8182	0.8448
12	0.7931	0.8270	0.8384	0.9103
13	0.7931	0.8374	0.9092	0.8648
14	0.8259	0.8846	0.9083	**0.9325**
15	0.7586	0.8470	0.8583	0.9203
16	0.7586	0.8346	0.8182	0.8876
17	0.7931	0.8250	0.8495	0.9031
18	0.7759	0.8225	0.8595	0.9176
19	0.7414	0.8236	0.8692	0.8276
20	0.7931	0.8146	0.8595	0.8276

While feature ranking techniques can be valuable for identifying the most relevant features in a dataset, their omission in our tool-based approach was deliberate and justified for several reasons. Firstly, the feature extraction methods employed, such as LIWC, BERT, Mediapipe, and DeepFace, are already designed to extract meaningful and discriminative features from textual and image data. These methods leverage sophisticated algorithms and pretrained models that inherently prioritize important features for classification tasks. Secondly, the use of PSO for feature selection further enhances the relevance of the extracted features by iteratively refining the feature subset based on their predictive performance. PSO optimized feature selection by evaluating the fitness of different feature subsets, ensuring that only the most informative features are retained for model training. Additionally, the high accuracy achieved by our GSE prediction model without feature ranking suggests that the selected features sufficiently capture the relevant information needed for accurate classification. Therefore, while feature ranking techniques can be beneficial in certain scenarios, our approach leverages the effectiveness of the feature extraction methods and PSO-based feature selection to achieve robust performance without the need for additional feature ranking.

C. CLASSIFICATION MODELS

In this section, we build classification models by applying different classification algorithms. We randomly divide our dataset of 435 Facebook users into training and testing sets, with a split of 70% for training and 30% for testing. Then, we apply conventional classifiers (i.e., Random Forest (RF), Xgboost, Adaptive Boosting (AdaBoost) [15], Stacking Ensemble technique) to predict the GSE score. Table 7 provides a comparative analysis of different classification models based on different feature extraction techniques: text, image, and hybrid (combining text and image features) for the tool-based feature extraction prediction model. The accuracy, F1-score macro, recall, and precision are reported for each combination of feature extraction and classification model.

D. RESULT FOR TOOL-BASED FEATURE EXTRACTION METHOD

In this section, we present the outcomes of experiments derived from a tool-based feature extraction approach across three distinct modalities: *i) text*, *ii) image*, and *iii) hybrid* datasets.

Table 7 shows the performance of different classification models based on tool-based extraction method. In the text data type, the LIWC+BERT feature extraction technique shows moderate accuracy across all classification models, ranging from 67.26% to 85.80%. The ensemble model (Stacking) achieves the highest accuracy, F1-score macro, and recall values. However, the precision of the models is relatively low. In the image data type, the combination of Mediapipe and DeepFace feature extraction technique leads to higher accuracy, ranging from 79.30% to 91.06%. The ensemble model (Stacking) consistently outperforms other models in terms of accuracy, F1-score macro, recall, and precision. In the hybrid type (text+image features), the accuracy further improves compared to the text and image types independently. The Random Forest (RF), Xgboost, and AdaBoost models achieve accuracy ranging from 81.32% to 93.25%. Once again, the ensemble model (Stacking) demonstrates the highest performance (accuracy-93.25%) across all evaluation metrics. Overall, the results indicate that combining both text and image features (hybrid) generally leads to better accuracy compared to using text or image features independently. However, our model emphasizes the effectiveness of combining multiple models to enhance the prediction performance. Additionally, it is worth noting that the precision values are generally lower than the recall values, indicating a higher rate of false positives in the predictions. Considering these findings, the hybrid feature extraction technique combined with the ensemble (Stacking) classification model yields the best overall performance, achieving high accuracy, F1-score macro, recall, and precision. Figure 6 displays the test accuracy curves for different machine learning models trained on the feature set obtained from a tool-based feature extraction approach.

V. DEEP LEARNING-BASED FEATURE EXTRACTION PREDICTION MODEL

We also apply deep learning-based feature extraction in developing our prediction model. There are numerous frameworks and necessary tools for constructing deep

TABLE 7. Comparison of the class performance from the tool-based feature selection method.

Data type	Feature extraction	Classification model	Accuracy	macro-F1	Recall	Precision
Text	LIWC + BERT (L_B)	Random Forest (RF)	67.26%	44.56	49.10	66.23
		Xgboost	71.10%	46.52	51.30	70.54
		AdaBoost	68.30%	39.86	49.86	66.89
		Stacking	**85.80%**	55.68	61.50	82.32
Image	Mediapipe + DeepFace (M_D)	Random Forest (RF)	83.57%	56.42	64.32	80.4
		Xgboost	88.10%	51.40	62.78	86.45
		AdaBoost	79.30%	46.20	52.10	78.64
		Stacking	**91.06%**	61.84	65.40	90.50
Hybrid	M_D+L_B	Random Forest (RF)	84.26%	51.80	60.54	83.79
		Xgboost	87.10%	51.00	61.20	86.47
		AdaBoost	81.32%	48.00	58.22	80.74
		Stacking	**93.25%**	53.67	62.45	91.89

learning models. Among these, the most popular framework is Keras.[5] Note that Tensorflow[6] is integrated into the Keras backbone, which supports the CPU/GPU environment. In this investigation, Keras and Tensorflow are executed in the GPU. Deep learning experiments are carried out using a computer equipped with the following components: an NVIDIA GeForce RTX 6 GB GPU, an Intel Core i7-12th processor, 24 GB of RAM, and an SSD hard disk. Furthermore, Google Colab[7] is employed for experiments when computer hardware is insufficient. Each algorithm is performed five times, with the highest average value recorded each time.

Deep learning models have the capability to automatically extract and learn features from data, making them well-suited for detecting qualities and characteristics in various domains. In the context of detecting aspects of human behavior (i.e., personality, values, sentiment, etc.), deep learning models offer significant advantages [55]. By analyzing both text and image data, these models are likely to identify patterns and cues that indicate the presence of self-efficacy. For example, the model can learn to recognize specific linguistic expressions or visual attributes associated with confidence in the text and images. The holistic approach combining text and image data enhances the accuracy of detecting self-efficacy by leveraging a comprehensive understanding of both modalities. Several studies [56], [57] also show improved performance when multi-modal content are used.

This section provides a detailed explanation of the proposed model's pipeline. Figure 5 depicts the general architecture of our deep learning-based feature extraction model. The model portion has been divided into three parts: 1) feature extraction, 2) feature fusion, and 3) training.

A. FEATURE EXTRACTION

We describe the feature extraction process by using deep learning based approach from text and image data in this part.

1) TEXT DATA

We apply two processes to extract features for the deep learning-based model. First, we employ contextual analysis for text data utilizing BERT embedding sentence-Transformers that we previously used in our tool-based architecture in Section IV. After that, we extract text features by using 1D-CNN, by applying convolutional operations on the text data which can effectively capture local patterns and dependencies within the text. These features allow us to extract meaningful features for various text-based tasks. First of all, we have taken a vector v^t from BERT feature extraction techniques. We pass the vector v^t through 1D-CNN at every stage:

$$h^{d,t} = tanh(W^d v^t + b^d) \tag{5}$$

$$h^d = [h^{d,1}, h^{d,2}, h^{d,3}, h^{d,4}, , , h^{d,t}] \tag{6}$$

$$p^d = Max^t[h^{d,1}, h^{d,2}, h^{d,3}, h^{d,4}, , , h^{d,t}] \tag{7}$$

$$H = [p1, p2, p3] \tag{8}$$

where W^d is the trainable weight matrix, b^d is the biases, and v^t is the embedded vector of the sentences. For a 1D-CNN, h^d stands for the corresponding hidden layers. The semantics of words in the context of the full status is encoded by H, wherein unit number of each 1D-CNN is u. The model performs better because we represent the query using all the hidden layers of the 1D-CNN as opposed to the final hidden layer.

2) IMAGE DATA

Computer vision and bio-medical engineering (i.e., segmenting lesions or anomalies) require a suitable segmentation technique for higher accuracy during classification tasks. In natural photos, it is also important to use precise segmentation to increase models' performance. However, minor segmentation mistakes in images might negatively impact user experience during model prediction in critical situations. As a result, we adopt UNet++, a new segmentation framework based on layered and dense hidden neurons. The main concept is to gradually enhance the high-dimensional maps from the encoder and then fuse them with semantically rich features from the decoder network, enabling the model to

[5] https://keras.io/
[6] https://www.tensorflow.org/
[7] https://colab.research.google.com/

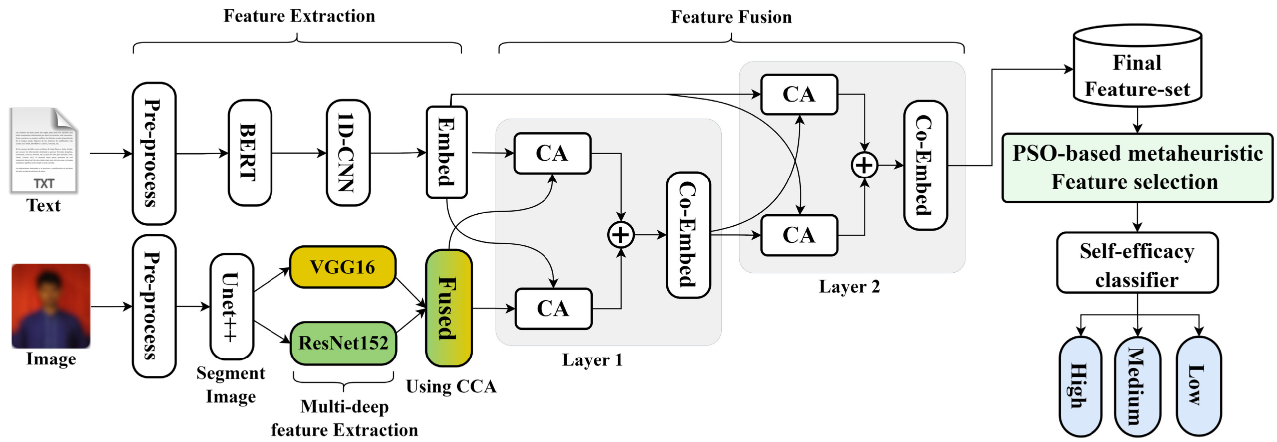


FIGURE 5. Architecture of our GSE prediction model from DL based feature selection method (CA and CCA represent co-attention and canonical correlation analysis, respectively).

capture fine-grained details of the original image effectively. Whenever the feature mappings from the decoder-encoder layers are semantically comparable, the network would handle a simpler learning task. High-resolution feature maps are swiftly forwarded from the decoder-encoder layers, resulting in the merging of feature maps with different semantic properties.

We have trained UNet++ model on OCHuman dataset [58] which is designed for studying heavily occluded humans and provides comprehensive annotations such as bounding boxes, human poses, and instance masks. After getting the segmented image from UNet++, we use deep learning-based multi-deep feature extraction using two different CNN-based pre-trained models VGG16 and ResNet-152.

Image Feature Extraction Using VGG16 and ResNet-152: We extract image features by using both VGG16 and ResNet-152 models. The VGG16 [59] model is used to extract deep features from users' profile photos. It employs a CNN and is capable of extracting 1,000 features from input images. The architecture consists of three convolutional layers, one pooling layer, and 16 regions. The features obtained from VGG16 are denoted as f_z^{VGG}, where z represents the feature index ranging from 1 to z. ResNet-152 [60] is also used to extract deep features from input users' profile photos. The model incorporates skip connections and performs well in image analysis tasks. The architecture includes identity mapping, convolution, ReLU, and batch normalization regions. The extracted features are represented as f_z^{Resnet}, where z denotes the feature index.

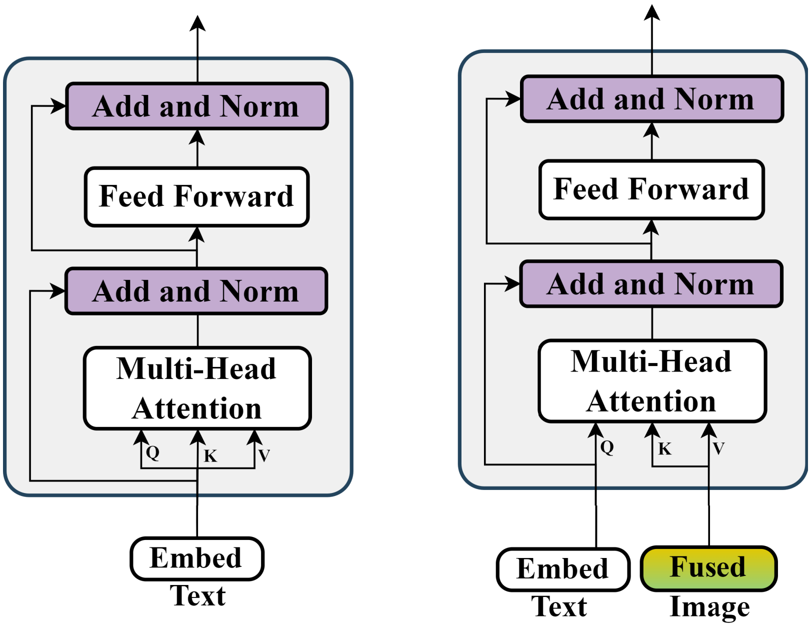
B. FEATURE FUSION METHOD

First of all, we fuse image features that we have taken from VGG16 and ResNet-152. VGG16 is suitable for datasets with simple and uniform visual patterns, while ResNet-152, with its skip connections and residual blocks, is better suited for datasets with complex visual patterns, variations in pose, lighting, or background. VGG16 is used to extract global image features, focusing on overall information without deep details, while ResNet-152 captures intricate and fine-grained details in images. Therefore, by using both of the techniques, we utilize the strengths of each model and enhance the overall feature representation for a more comprehensive understanding of the profile photos. A few studies [61] also apply similar technique in their experiments.

To fuse those features, we have used CCA because it facilitates the identification of shared information, the creation of a unified representation, and the assessment of the fusion quality. Several studies [14], [62] also use modified CCA to fuse features and get excellent results. After obtaining the fused image feature using CCA, we apply another fusion technique called the co-attention mechanism. This mechanism is typically used for multimodal data that combines both text and image information. In our case, we have already obtained the final text feature using a 1D-CNN and the image feature using CCA. The co-attention mechanism is then employed to further integrating these features and enhancing the fusion process.

1) MERGING IMAGE AND TEXT FEATURES USING COATTENTION

Motivated by several studies [63], [64], we incorporate attention technique into our classification model, particularly for handling multimodal data. Vaswani et al. [65] propose a multi-head self-attention (MSA) which focused on single-type data for attention and can capture the overall relationships and dependencies among all positions within a text sequence. Since, our data consists of multiple modalities (i.e., text and image), we applied the modified MSA introduced by Yang et al. [66]. The approach can be simulated as if a psychologist is assessing a patient's GSE score by first observing their facial expressions and then reading their written text. In our case, we consider a similar technique, where we fuse multimodal data by using a coattention-based approach. This allows us to effectively combine information



(a) MSA block

(b) CA block

FIGURE 6. **(a) Represents Multi-head self-attention (MSA) block and (b) represents Co-attention (CA) block.**

from both text (represented by users' Facebook statuses) and images (represented by user profile photos) in our dataset.

The co-attention (CA) block is extended from the MSA block, as shown in Figure 6 (a) and (b). For a CA block, the queries are from one modality while keys and values are from another modality. In particular, the query matrix is used as a residual item after the multi-head attention sublayer. The rest of the architecture is the same as MSA. The CA block produces an attention-pooled feature for one modality conditioned on another modality. If Q comes from the text and K and V comes from the attached image, the attention value calculated using Q and K can be used as a measure of the similarity between the text and image and then assigns weight of the image, where the query (Q) represents the information of interest, the key (K) provides context, and the value (V) contains the actual information. Just like humans, after reading the text, they pay more attention to the areas in the image that are similar to the text. Our co-attention unit simulates this process and learns inter-dependencies between different features among the content with different modalities.

2) CO-ATTENTION FUSION (CAF) LAYER

We obtain a Co-attention Fusion (CAF) layer by connecting two CA blocks in parallel, as shown in Figure 5. Given two CA blocks with different features, the CAF layer computes queries, keys, and values for each CA block as in an MSA block. Then the keys and values of one CA block are passed as input to another CA block. The outputs of two CA blocks are concatenated together and then fed into a fully connected layer to get the fused representation. The CAF layer models dense interactions between input modalities by exchanging their information. In order to fuse multimodal features deeply, we apply two consecutive CAF layers in a stacked manner. In the first CAF layer, we combine image and text features, resulting in a feature representation denoted as RC^1. To further enrich this representation, we introduce another attention mechanism using text features on the RC^1 representation, resulting in a more informative feature representation termed RC^2. The fusion process is progressive, and the output of each CAF layer is one of the inputs of the next layer (see Figure 5). We first fuse spatial domain representation R_S and text embedding-domain representation R_T in the first CAF layer and obtain RC^1. Then RT is enhanced to fuse with RC^1 in the second CAF layer which outputs RC^2. The output vector of each CAF layer is d-dimensional. The calculation processes are formulated as follows.

$$R_{C_{S<--T}} = R_S + MHA(R_S, R_T, R_T) \tag{9}$$

$$RC'_{S<--T} = R_{C_{S<--T}} + FFN(R_{C_{S<--T}}) \tag{10}$$

$$R_{C_{T<--S}} = R_T + MHA(R_T, R_S, R_S) \tag{11}$$

$$RC'_{T<--S} = R_{C_{T<--S}} + FFN(R_{C_{T<--S}}) \tag{12}$$

$$RC^1 = (R_{C_{S<--T}} \oplus R_{C_{T<--S}}).WC^1 \tag{13}$$

where $RC'_{S<--T}\epsilon\mathbb{R}^d$ is the attention-pooled feature for the spatial domain conditioned on the text domain, $RC'_{T<--S}\epsilon\mathbb{R}^d$ is the attention-pooled feature for the text domain conditioned on the spatial domain, and $WC^1\epsilon\mathbb{R}^{2dd}$

is the projection matrix of the first CAF layer. RC^1 is transformed to be a $(dx1)$-dimensional representation before being input to the next CAF layer. Specifically, the first and the third CAF layers share parameters, and the second and the fourth CAF layers share parameters.

Finally, we have obtained the fused multimodal feature representation RC^2, which incorporates features from the text and spatial domain. This fused representation, denoted as $X = RC^2$, is utilized for prediction purposes.

C. FEATURE SELECTION

We have used the same feature selection technique used in the tool-based model in section IV-B because, from both feature extraction models, we have a set of features and need to identify which features are important to train for final prediction. In Table 8, the performance metrics of different machine learning algorithms are presented for various feature sets (1 to 20). Notably, feature set 9 consistently demonstrates higher accuracy across multiple algorithms. Actually, we utilized the PSO method to select 237 features, which resulted in the highest performance.

The underlying capabilities of deep neural networks led to the decision to skip feature ranking in order to estimate self-efficacy from user images and Facebook status using deep learning. Rather than requiring explicit feature ranking procedures, these models may automatically learn and extract significant features from unprocessed data. By using the complete dataset, including potentially useful but obscure patterns, deep learning reduces information loss, avoids human feature engineering, and catches complex relationships. Moreover, feature ranking is not so important in this situation because deep learning models can manage complex, non-linear connections and are resilient to irrelevant features.

TABLE 8. PSO based fitness score for various machine learning models depending on the different feature sets from DL-based features.

Fetaure set	Random Forest (RF) fitness	Xgboost fitness	AdaBoost fitness	Stacking fitness
1	0.7356	0.7659	0.7434	0.7276
2	0.6724	0.7679	0.7513	0.7375
3	0.7123	0.7719	0.7112	0.7046
4	0.7213	0.7556	0.7414	0.7016
5	0.6823	0.7355	0.7510	0.7126
6	0.7237	0.7469	0.7418	0.7245
7	0.6913	0.7557	0.7516	0.7575
8	0.6924	0.7458	0.7410	0.7086
9	0.6834	0.7759	0.7316	**0.8187**
10	0.7392	0.7669	0.7116	0.7976
11	0.7182	0.7452	0.7424	0.7960
12	0.7374	0.7121	0.7315	0.7270
13	0.7247	0.7251	0.7416	0.8174
14	0.6824	0.7550	0.7714	0.8046
15	0.7274	0.7557	0.7234	0.7470
16	0.7293	0.7857	0.7210	0.8046
17	0.6928	0.7453	0.7510	0.8150
18	0.7493	0.7743	0.7604	0.8125
19	0.7593	0.7652	0.7484	0.8036
20	0.7392	0.7764	0.7714	0.8146

D. CLASSIFICATION MODELS

In this section, we construct classification models using various classification algorithms. Our dataset, consisting of 435 Facebook users, is randomly divided into training and testing datasets with a split of 70% and 30% respectively. We apply conventional classifiers, including Random Forest (RF), Xgboost, Adaptive Boosting (AdaBoost), and Stacking Ensemble technique, to predict the GSE score. Table 9 presents a comparative analysis of different classification models based on three feature extraction techniques: text, image, and a hybrid approach combining text and image features. The table includes accuracy, F1-score macro, F1-score, recall, and precision metrics for each combination of feature extraction and classification model.

E. RESULT FOR DL-BASED FEATURE EXTRACTION METHOD

In this section, we present the results of experiments conducted using a feature extraction approach based on deep learning across three distinct modalities: i) text, ii) image, and iii) hybrid datasets.

Table 9 provides a comparative analysis of different classification models based on DL based feature extraction methods for three types of data: text, image, and hybrid (combining text and image features). The accuracy, F1-score macro, F1-score, recall, and precision are reported for each combination of feature extraction and classification model. In the text type, the BERT + 1D-CNN feature extraction technique achieves moderate accuracy, ranging from 57.30% to 64.80%. The ensemble model (stacking) demonstrates the highest F1-score macro, F1-score, recall, and precision among all the models. In the image type, utilizing VGG16 and ResNet-152 for feature extraction results in accuracy ranging from 68.15% to 73.06%. The ensemble model (stacking) consistently outperforms other models in terms of accuracy, F1-score macro, recall, and precision. In the hybrid type (text+image), the accuracy improves compared to the text type. The co-attention feature extraction technique achieves accuracy ranging from 75.10% to 81.87%. Similar to the other types, the ensemble model (stacking) demonstrates superior performance across all evaluation metrics. Overall, the results suggest that combining text and image features (hybrid) generally leads to better accuracy compared to using text or image features independently. The ensemble model consistently performs well across all feature extraction techniques and achieves the highest accuracy, F1-score macro, recall, and precision values. This highlights the effectiveness of combining multiple feature set for improved prediction performance. In terms of feature extraction techniques, the co-attention method in the hybrid type shows the highest accuracy (81.87%) and F1-score macro among the three types of data.

The comparative analysis of the ablation study for DL-based feature extraction method presented in Table 10 elucidates the impact of various configurations on the GSE prediction model. The baseline model, incorporating BERT for text, UNet++ with VGG16/ResNet-152 for

TABLE 9. Comparison of the performance of the DL models.

Data type	Feature extraction	Classification model	Accuracy	macro-F1	Recall	Precision
Text	BERT(L_B)+1D-CNN	Random Forest (RF)	64.36%	47.36	55.10	63.75
		Xgboost	57.30%	46.54	57.30	56.41
		AdaBoost	63.30%	47.86	56.86	61.57
		Ensemble (Stacking)	**64.80%**	48.68	58.50	63.50
Image	VGG16 + ResNet-152	Random Forest (RF)	71.50%	63.42	69.32	71.35
		Xgboost	68.15%	57.40	67.78	66.48
		AdaBoost	72.35%	53.20	58.10	70.54
		Ensemble (Stacking)	**73.06%**	68.84	71.40	72.54
Hybrid	Co-attention	Random Forest (RF)	78.26%	57.80	66.54	76.89
		Xgboost	75.10%	57.00	67.20	74.85
		AdaBoost	77.42%	52.00	64.22	75.84
		Ensemble (Stacking)	**81.87%**	59.67	68.45	80.75

TABLE 10. Ablation study experiment configurations and components.

Experiment	Text Feature Extraction	Image Feature Extraction	Feature Fusion	Feature Selection	Classification Models	Time Complexity	Accuracy
Selected (Full Model)	BERT + 1D-CNN	UNet++ + VGG16/ ResNet-152	CCA + Co-attention	PSO	Random Forest (RF), Xgboost, AdaBoost, Stacking Ensemble	58.06 M	**81.87%**
Text Feature Extraction only (BERT)	BERT Only	UNet++ + VGG16/ ResNet-152	CCA/Co-attention	PSO	Random Forest (RF), Xgboost, AdaBoost, Stacking Ensemble	45.25 M	80.72%
Image Feature Extraction only (VGG16/ ResNet-152)	BERT + 1D-CNN	VGG16/ ResNet-152 Only	CCA/Co-attention	PSO	Random Forest (RF), Xgboost, AdaBoost, Stacking Ensemble	51.01 M	78.38%
CCA Feature Fusion only	BERT + 1D-CNN	UNet++ + VGG16/ ResNet-152	CCA Only	PSO	Random Forest (RF), Xgboost, AdaBoost, Stacking Ensemble	48.26 M	80.01%
Co-attention Feature Fusion only	BERT + 1D-CNN	UNet++ + VGG16/ ResNet-152	Co-attention Only	PSO	Random Forest (RF), Xgboost, AdaBoost, Stacking Ensemble	50.46 M	78.12%
Without Feature Selection	BERT + 1D-CNN	UNet++ + VGG16/ ResNet-152	CCA + Co-attention	No Feature Selection	Random Forest (RF), Xgboost, AdaBoost, Stacking Ensemble	48.20 M	78.06%
Individual Classification Algorithms	BERT + 1D-CNN	UNet++ + VGG16/ ResNet-152	CCA + Co-attention	PSO	Random Forest (RF) Only / Xgboost Only / AdaBoost Only	50.22 M	76.86% - 80.25%

image, CCA with co-attention for feature fusion, PSO for feature selection, and diverse classifiers, achieves the highest accuracy at 81.87%, albeit with a relatively higher time complexity of 58.06 million (M). Streamlining the approach by focusing solely on BERT for text feature extraction results in a marginally reduced accuracy of 80.72%, accompanied by a decreased time complexity of 45.25 M. Conversely, relying on VGG16/ResNet-152 exclusively for image feature extraction lowers accuracy further to 78.3%. Emphasizing the importance of feature fusion, the exclusive use of CCA maintains high accuracy (80.01%) with a reduced time complexity of 48.26 M, while relying solely on co-attention yields an accuracy of 78.12%. Notably, eliminating feature selection slightly diminishes accuracy to 78.06%. Implementing individual classification algorithms, such as Random Forest (RF) only, Xgboost only, and AdaBoost only, results in the lowest accuracy of 76.86%. This detailed analysis underscores the nuanced trade-offs between accuracy and computational complexity within each configuration, providing valuable insights for model optimization.

Table 11 and Figure 7 present a comparative analysis of various CNN models used for image feature extraction. Each row shows different CNN architectures, including VGG16, VGG19, ResNet-50, ResNet-152, Inception V3, and MobileNetV2. Information provided includes the number of layers, parameter counts, training epochs, time taken per epoch, optimizer used (in this case, *Adam*), batch sizes, input image sizes, and output feature map dimensions. This comparison enables a comprehensive understanding of the architecture complexity, parameter sizes, training times, and feature map sizes of these CNN models, aiding in selecting an appropriate model for image feature extraction tasks.

VI. COMPARING THE OUTCOME OF DL-BASED AND TOOL-BASED METHODS

In this section, we have contrasted the outcomes of experiments conducted with tool-based and DL-based approaches across three specific modalities: i) text, ii) image, and iii) hybrid datasets.

Table 13 presents a comprehensive comparison of the methodologies utilized for GSE prediction, categorized into

TABLE 11. Different parameters of CNN models.

Model to extract image features	No. of layers	No. of Parameters	Epochs	Time per Epoch	Optimizer	Batch Size	Image size	Feature size
VGG16	16	138.4 M	100	141-143s	Adam	32	224x224	7x7x512
VGG19	19	143.7 M	100	142-145s	Adam	32	224x224	7x7x512
ResNet-50	50	25.6 M	100	148-150s	Adam	32	224x224	7x7x2048
ResNet-152	150	115.6 M	100	151-153s	Adam	32	224x224	7x7x2048
Inception V3	48	23.9 M	100	146-148s	Adam	32	224x224	8x8x2048
MobileNetV2	53	3.5 M	100	150-151s	Adam	32	224x224	7x7x1280

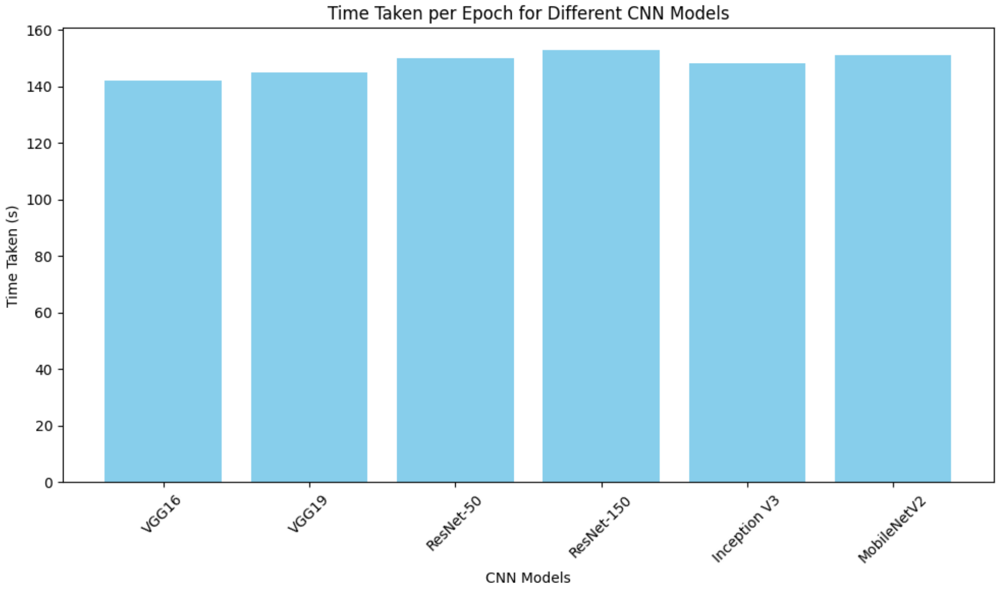


FIGURE 7. Visualization depicting the time required per epoch for distinct CNN models.

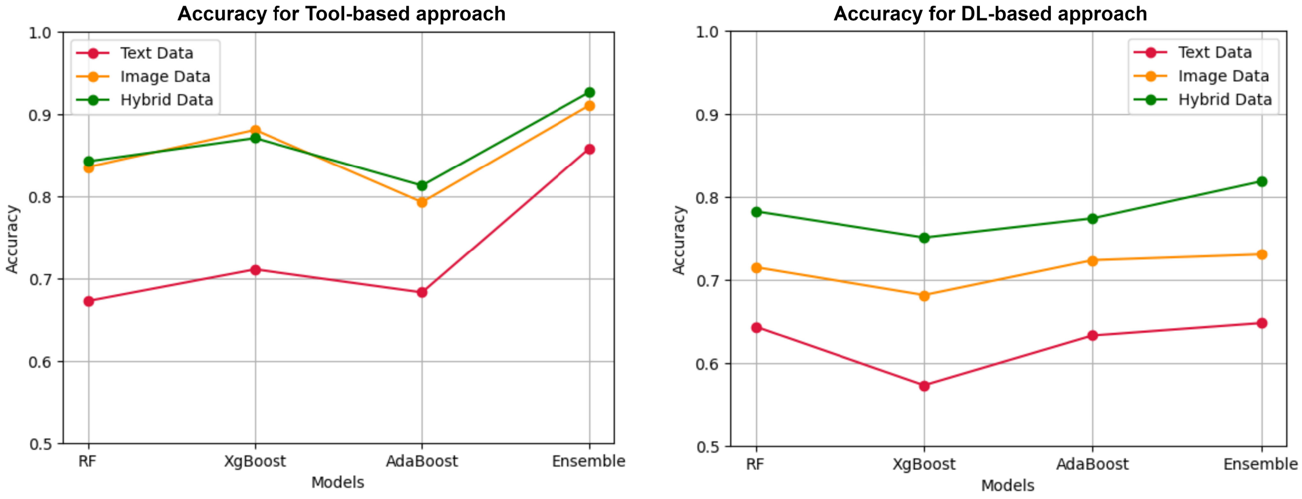


FIGURE 8. Test accuracy curves from Tool-based and DL-based feature extraction models.

tool-based and deep learning-based approaches. In the tool-based approach, feature extraction methods such as LIWC (word count, sentiment analysis), BERT, Mediapipe (keypoints extraction, face recognition), and DeepFace (VGG-Face model for emotion recognition) were employed, coupled with traditional machine learning models including Random Forest (RF), Xgboost, and AdaBoost. Notably, the tool-based approach achieved an impressive accuracy

TABLE 12. Comparison among the best performing models.

Model type	Data type	Best Accuracy	F1-score	macro-F1	Recall	Precision
Tool based	Text	85.80%	71.30	55.68	61.50	82.32
	Image	91.06%	82.45	61.84	65.40	90.50
	Hybrid	**93.25%**	81.50	53.67	62.45	91.89
Deep Learning	Text	64.80%	57.98	48.68	58.50	63.50
	Image	73.06%	63.13	68.84	71.40	72.54
	Hybrid	81.87%	64.53	59.67	68.45	80.75

of 93.25%. Conversely, the deep learning-based approach utilized feature extraction techniques like BERT, 1D-CNN, UNet++, VGG16, and ResNet-152, alongside similar machine learning models as in the tool-based approach. However, it exhibited a slightly lower accuracy of 81.87%. This discrepancy in accuracy underscores the effectiveness of combining various feature extraction methods in the tool-based approach, highlighting its superiority in GSE prediction compared to the deep learning-based approach. Figure 8 illustrates the test accuracy curves for various machine learning models that were trained using the feature set extracted through a Tool-based and DL-based feature extraction approach.

Table 12 provides a comparison of different models based on their performance in terms of accuracy, F1-score macro, F1-score, recall, and precision. The models are categorized into two types: tool-based and deep learning-based depending on their feature extraction method. Under the tool-based category, three data types were evaluated: text, image, and a hybrid of both. The best accuracy was achieved by the hybrid model, with a score of 92.66%. The deep learning-based models also assessed text, image, and hybrid data types. In this category, the hybrid model shows the highest accuracy of 81.87%. These results highlight the superior performance of the hybrid models in both the tool-based and deep learning-based approaches, particularly when incorporating both image and text data together.

This difference in performance can be attributed to the inherent characteristics of the two approaches (tool-based and DL based feature extraction). The tool-based approach relies on explicitly defined features extracted from text and image data using established tools and algorithms. It benefits from the interpretability and domain-specific knowledge embedded in these tools, leading to a strong predictive performance. On the other hand, while deep learning models excel in capturing intricate patterns and representations in complex data, they can be more challenging to interpret and may require larger amounts of labeled data for training. We obtained a weaker DL based method, where other coattention approaches have achieved high accuracy with multimodal data due to several factors. An uneven composition of the dataset may increase diversity between image and text data which eventually limits the ability to capture the relationships between the image and text modalities. Image and text data inconsistencies may also hinder the performance of the coattention model. In this particular case, the tool-based approach, with its explicit feature extraction and strong predictive performance, may outperform the deep learning-based approach.

VII. DISCUSSION

Our study is the first research that predicts users' self-efficacy from their Facebook-based multiple interaction modalities, i.e., statuses and profile photos. We observe that different levels of a user's self-efficacy may have different impacts on their personality and lifestyle activities.

A. CORRELATIONS WITH PERSONALITY TRAITS

Personality traits might have correlations with users' self-efficacy scores. Khan et al. [22] find correlations between Big5 personality traits and LIWC scores. The relevant LIWC categories are: *work [67], tentative [68], positive emotion [48], and negative emotion [69]*. Similarly, in our study, we also find a few traits have also correlations between self-efficacy and LIWC scores (Section IV) which include work [70], tentative [71], time [72], positive emotion [55], negative emotion [69] and affect [73]. We observe that a user with a high GSE score likely has positive correlations with *tentative* and *positive emotion* LIWC scores. Hayat et al. [74] find that students' learning related emotions (i.e., positive and negative) are strongly connected with their self-efficacy after conducting an experiment among 279 medical students. Actually, emotions influence the meta-cognitive learning processes, which in turn impact on performance of an individual in any context. Similarly other studies [75], [76] also corroborate our findings. In contrast, we also observe that a user with high *negative emotional* score has strong correlations with low GSE scores. Luszczynska et al. [38] also find similar insights in their socio-psychological study of GSE. We also find that an individual with low GSE score tends to demonstrate high usage of negative pattern of words such as *sadness, anger* and *anxiety* in her Facebook statuses. Caprara et al. [77] show that anger, sadness, fear and shame are associated negatively with users' self-efficacy. They mention that these attributes do not influence an individual much to quest for life satisfaction. Motro et al. [78] also made the same conclusion from a face-based emotion recognition software and their self-efficacy scores of 96 participants.

However, in our PSO based feature selection step, we also find a few selected LIWC features to predict GSE scores which are not explainable in real life situations such as *article* and *six letter* categories. Figure 9 represents how LIWC

TABLE 13. Feature extraction and machine learning model details for GSE prediction.

Model Approach	Feature Extraction Methods	Feature Extraction Parameters	Machine Learning Models	ML Model Parameters	GSE Score (Accuracy)
Tool-based Approach	LIWC, BERT, Mediapipe, DeepFace	LIWC (word count, sentiment), BERT (max_length=128, layers=12, attention_heads=8), Mediapipe (keypoints: full, face recognition), DeepFace (model=VGG-Face, emotion recognition)	Traditional ML Models (Random Forest (RF), Xgboost, AdaBoost)	Random Forest (RF) (n_estimators=150, max_depth=12), Xgboost (n_estimators=200, learning_rate=0.05, max_depth=7) Adaboost (n_estimators=100, learning_rate=0.1)	93.25%
Deep Learning-based Approach	BERT, 1D-CNN, UNet++, VGG16, ResNet-152, CCA, Co-attention	BERT (max_length=256, layers=6, attention_heads=12), 1D-CNN (kernel_size=5), UNet++ (backbone-152), VGG16 (pooling=avg), ResNet-152 (pooling=avg), CCA (components=120), Co-attention (heads=6)	Traditional ML Models (Random Forest (RF), Xgboost, AdaBoost, Stacking)	Random Forest (RF) (n_estimators=250, max_depth=15), Xgboost (n_estimators=300, learning_rate=0.1, max_depth=8), AdaBoost (n_estimators=150, learning_rate=0.05)	81.87%

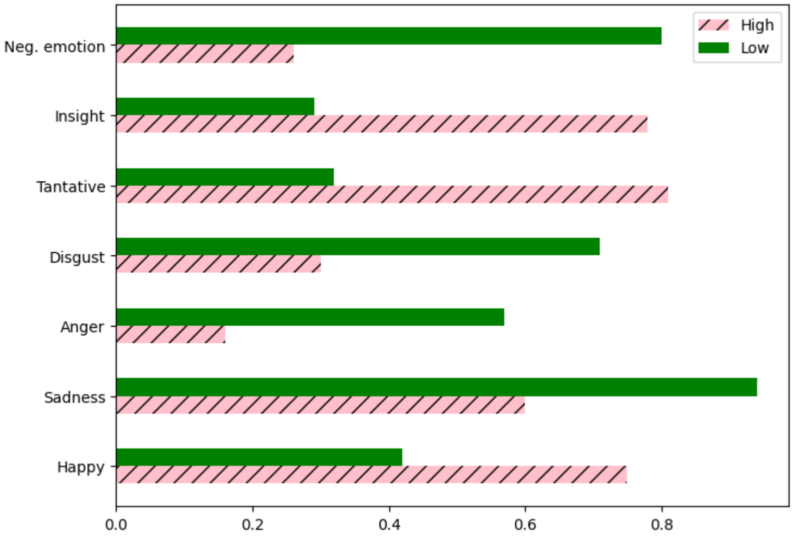


FIGURE 9. Correlation between personality attributes and GSE.

scores (i.e., negative emotion, tentative, anger, sadness, joy, etc.) for deriving personality traits and GSE are correlated.

B. ASSOCIATION WITH IMAGE DATA

In our image dataset, we utilized *Mediapipe* and *DeepFace* libraries for extracting human facial landmarks and emotional attributes, respectively. The output of Mediapipe face mesh finds association between emotions based on facial expressions (emotion can be considered part of a person's *personality attribute* as it is a crucial aspect of their inner experience and has ties to their traits, actions, and attitudes.), physical attractiveness [79] based on facial symmetry, and age [80] by analyzing facial features. To observe whether image features significantly vary based on different self-efficacy levels (i.e., low, medium, and high), we randomly select profile photos of 30 users of each level and extracted feature maps from the 2D-CNN convolutional layer. We then test these features maps by using analysis of variance (ANOVA) [81] looking at statistics such as *standard deviation*, *perimeter area ratio*, *skewness*, and *kurtosis*. The results showed a critical value of ($\rho < 0.05$), which accept an alternate hypothesis. The critical value indicates that there is a significant difference among the facial expressions of users of different self-efficacy levels.

C. EMOTIONAL VARIABILITY IN OUR DATASET

Table 4 shows the statistics of different DeepFace attributes in our dataset. The features presented in the data exhibit diverse

TABLE 14. Research agenda on self-efficacy analysis.

Research agenda	Challenge/research question	Research directions
Analysis of audio-visual content	Can human audio content, even audio-visual content [82] reveal his self-efficacy?	By feature merging and using a co-attention mechanism of different content, we may deploy the model.
Analysis of human brain signals	Brain signals may be another determinant for predicting self-efficacy of an individual [83].	By analyzing EEG signals of an individual, we can apply attention and residual blocks combined with advanced LSTM and CNN models
Gait analysis	Can human kinetic parameters and musculoskeletal functions reveal human self-efficacy?	Since locomotive data is sequential, any RNN based model fusioning with a deep learning based model should work fine [84].
Explainability of the model	Can we visualize parameters and model performance in an AI model?	Image heatmaps, SHAP and other tools show the importance of features to predict self-efficacy from different content [85].
Consistency among different social networks/content types	Do different social media platforms and content types produce similar self-efficacy scores?	A comparative study in the same sample of subjects from different social networks, we apply similar ML and DL techniques and compare the scores.
Change of self-efficacy	Is self-efficacy a stable human attribute [86]?	By computing time series scores of human self-efficacy (using ARIMA, LSTM, etc.) and evaluating human influence from social media dynamics.
Association with other human attributes	Are Big5 traits, emotion, and sentiment associated with human self-efficacy?	For the same sample of subjects, we can compute the relationship between these attributes and find practical implications.

scores that reflect the variability and distribution of different emotions among individuals. For instance, the mean score of 5.95 for anger indicates a moderate presence, with a wide range of values from 0.001 to 99.59, highlighting significant variation in anger intensity. Conversely, disgust is represented by a minimal mean score of 0.00 which demonstrates a limited expression of this emotion. Fear, with a mean score of 11.03, shows a moderate presence accompanied by values ranging from 0.001 to 98.43, indicating variability in fear intensity. Happiness, with a mean score of 26.02, exhibits a relatively high presence, spanning from 0.001 to 100.00, and a wide standard deviation of 39.16, is suggesting substantial variation in happiness intensity. Similarly, sadness and surprise demonstrate moderate presence, with diverse scores ranging from 0.001 to 99.97. The neutral emotion, characterized by a mean score of 42.21, exhibits a relatively high presence, with values ranging from 0.001 to 100.00 and a notable standard deviation of 40.81, highlighting variability in neutrality expression. These diverse scores reflect the natural distribution in facial expressions and emotional responses observed among individuals, encompassing a wide spectrum of emotions from low GSE (disgust) to moderate GSE (fear and surprise) and high GSE (happiness and neutrality) scores.

D. LIMITATIONS AND FUTURE DIRECTIONS

Our study has several limitations. For example, Bandura [16] describes that four psychological dimensions such as: mastery experiences, vicarious experiences, verbal persuasion and emotional and physiological states which can influence GSE score of an individual. However, in our study we cannot compute these dimensional scores independently from their Facebook statuses and profile photos. If an individual is confident about her skills, then she is likely to achieve success. Her *mastery of experience* increases after more practices and performing tasks successfully [87] which can be mined by analyzing her statuses rigorously. By seeing people succeeding at a particular task, another individual gets confidence and performs well in that task which is called *vicarious experience*. Verbal persuasion and physiological and affective states another two important dimensions of users' self-efficacy. If we process users' Facebook statuses and comments carefully by using rigorous natural language processing (NLP) tasks, then we could predict users' resultant self-efficacy levels by integrating their above four dimensions. However, another major shortcomings of our paper is that we do not find any priority of content in predicting users' self-efficacy over another content. For example, a user may share her self-efficacy capability through her profile photos, while another user is proficient with her writing. Therefore, building a weighted machine learning model for predicting self-efficacy could be an interesting research topic. Furthermore, building an automated self-efficacy based prediction system from social media content may pose some ethical issues. For example, an inactive Facebook user who uses a random profile photo which may not represent his idealistic self-efficacy level. This wrong prediction may not recommend the user to grab a competitive job from an employer though he is a potential candidate for the role in real life.

To enhance this work, we have several future research avenues for the prospective researchers. Table 14 presents different future research directions.

VIII. CONCLUSION

Self-efficacy is the belief in one's capability to succeed in life. Self-efficacy has a significant impact on individual's life as it determines the setting of goals, levels of motivation, responses to stress, overall performance, and resilience, influencing their overall life path and well-being. In this study, we have explored the profound concept of self-efficacy, delving into its implications for individuals' lives. With a robust dataset collected through a gold-standard survey aligned with Bandura's theory, we have scrutinized textual and visual content from social media platform. Employing two distinct feature extraction methodologies, we have harnessed the power of both tool-based approaches, leveraging LIWC, BERT, Mediapipe, and DeepFace, and advanced deep learning techniques, incorporating BERT, 1D-CNN, UNet++, VGG16, ResNet-152, and co-attention models. Notably, our tool-based approach has demonstrated an outstanding predictive accuracy of 93.25%. Conversely, the deep learning-based approach has yielded a satisfactory accuracy of 81.87%. These findings elucidate nuanced connections between self-efficacy levels and social media content posting patterns, underscoring the potential for leveraging digital footprints to gauge individuals' psychological constructs.

REFERENCES

[1] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *Proc. Nat. Acad. Sci. USA*, vol. 110, no. 15, pp. 5802–5805, Apr. 2013.

[2] J. Chen, G. Hsieh, J. U. Mahmud, and J. Nichols, "Understanding individuals' personal values from social media word use," in *Proc. 17th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2014, pp. 405–414.

[3] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 5, 2011, pp. 450–453.

[4] A. Bandura, "Self-efficacy mechanism in human agency," *Amer. Psychologist*, vol. 37, no. 2, pp. 122–147, 1982.

[5] M. D. Back, J. M. Stopfer, S. Vazire, S. Gaddis, S. C. Schmukle, B. Egloff, and S. D. Gosling, "Facebook profiles reflect actual personality, not self-idealization," *Psychol. Sci.*, vol. 21, no. 3, pp. 372–374, Mar. 2010.

[6] D. Wang, L. Xu, and H. C. Chan, "Understanding users' continuance of Facebook: The role of general and specific computer self-efficacy," in *Proc. ICIS*, 2008, p. 168.

[7] D. R. Compeau and C. A. Higgins, "Computer self-efficacy: Development of a measure and initial test," *MIS Quart.*, vol. 19, no. 2, p. 189, Jun. 1995.

[8] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, gender, and age in the language of social media: The open-vocabulary approach," *PLoS ONE*, vol. 8, no. 9, Sep. 2013, Art. no. e73791.

[9] A. Eftekhar, C. Fullwood, and N. Morris, "Capturing personality from Facebook photos and photo-related activities: How much exposure do you need?" *Comput. Hum. Behav.*, vol. 37, pp. 162–170, Aug. 2014.

[10] J. W. Pennebaker, R. J. Booth, and M. E. Francis, "Linguistic inquiry and word count (LIWC2007)," Comput. Softw., Texas Tech Univ., Austin, TX, USA, Tech. Rep., 2007.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[12] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle swarm optimization: A comprehensive survey," *IEEE Access*, vol. 10, pp. 10031–10061, 2022.

[13] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Anested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support Conjunction with MICCAI*. Granada, Spain: Springer, 2018, pp. 3–11.

[14] F. Afza, M. Sharif, M. A. Khan, U. Tariq, H.-S. Yong, and J. Cha, "Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine," *Sensors*, vol. 22, no. 3, p. 799, Jan. 2022.

[15] A. Shahraki, M. Abbasi, and Ø. Haugen, "Boosting algorithms for network intrusion detection: A comparative evaluation of real AdaBoost, Gentle AdaBoost and modest AdaBoost," *Eng. Appl. Artif. Intell.*, vol. 94, Sep. 2020, Art. no. 103770.

[16] A. Bandura, "Self-efficacy: Toward a unifying theory of behavioral change," *Psychol. Rev.*, vol. 84, no. 2, pp. 191–215, 1977.

[17] D. H. Schunk and M. K. DiBenedetto, "Self-efficacy and human motivation," in *Advances in Motivation Science*, vol. 8. Amsterdam, The Netherlands: Elsevier, 2021, pp. 153–179.

[18] M. S. H. Mukta, M. E. Ali, and J. Mahmud, "User generated vs. supported contents: Which one can better predict basic human values?" in *Proc. Int. Conf. Social Informat. (SocInfo)* Bellevue, WA, USA: Springer, 2016, pp. 454–470.

[19] M. S. H. Mukta, M. E. Ali, and J. Mahmud, "Identifying and validating personality traits-based homophilies for an egocentric network," *Social Netw. Anal. Mining*, vol. 6, no. 1, pp. 1–16, Dec. 2016.

[20] M. S. H. Mukta, S. Islam, S. Shatabda, M. E. Ali, and A. Zaman, "Predicting academic performance: Analysis of students' mental health condition from social media interactions," *Behav. Sci.*, vol. 12, no. 4, p. 87, Mar. 2022.

[21] A. S. Sakib, M. S. H. Mukta, F. R. Huda, A. K. M. N. Islam, T. Islam, and M. E. Ali, "Identifying insomnia from social media posts: Psycholinguistic analyses of user tweets," *J. Med. Internet Res.*, vol. 23, no. 12, Dec. 2021, Art. no. e27613.

[22] E. M. Khan, M. S. H. Mukta, M. E. Ali, and J. Mahmud, "Predicting users' movie preference and rating behavior from personality and values," *ACM Trans. Interact. Intell. Syst.*, vol. 10, no. 3, pp. 1–25, Sep. 2020.

[23] J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," in *Proc. CHI Extended Abstr. Human Factors Comput. Syst.*, May 2011, pp. 253–262.

[24] G. Hsieh, J. Chen, J. U. Mahmud, and J. Nichols, "You read what you value: Understanding personal values and reading interests," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, Apr. 2014, pp. 983–986.

[25] F. Celli, E. Bruni, and B. Lepri, "Automatic personality and interaction style recognition from Facebook profile pictures," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 1101–1104.

[26] L. Liu, D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. Ungar, "Analyzing personality through social media profile picture choice," in *Proc. 10th Int. AAAI Conf. Web Social Media*, 2016, pp. 211–220.

[27] C. Segalin, F. Celli, L. Polonio, M. Kosinski, D. Stillwell, N. Sebe, M. Cristani, and B. Lepri, "What your Facebook profile picture reveals about your personality," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 460–468.

[28] H. Yildiz, "Prediction of pre-service teachers' academic self-efficacy through machine learning approaches," *Afr. Educ. Res. J.*, vol. 11, no. 1, pp. 32–44, Jan. 2023.

[29] K. Karatağ, I. Arpaci, and S. Süer, "Predicting academic self-efficacy based on self-directed learning and future time perspective," *Psychol. Rep.*, Jul. 2023, Art. no. 00332941231191721.

[30] B. Tan and M. Cutumisu, "Employing tree-based algorithms to predict students' self-efficacy in PISA 2018," in *Proc. 15th Int. Conf. Educ. Data Mining*, 2018, p. 634.

[31] M. Jamjoom, E. Alabdulkreem, M. Hadjouni, F. Karim, and M. Qarh, "Early prediction for at-risk students in an introductory programming course based on student self-efficacy," *Informatica*, vol. 45, no. 6, pp. 1–9, Aug. 2021.

[32] A. Bandura and S. Wessels, "Self-efficacy," in *Encyclopedia of Human Behavior*. New York, NY, USA: Academic, 1994.

[33] A. Bandura, "Reflections on self-efficacy," *Adv. Behaviour Res. Therapy*, vol. 1, no. 4, pp. 237–269, Jan. 1978.

[34] A. Bandura, "The explanatory and predictive scope of self-efficacy theory," *J. Social Clin. Psychol.*, vol. 4, no. 3, pp. 359–373, Sep. 1986.

[35] G. M. Marakas, M. Y. Yi, and R. D. Johnson, "The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research," *Inf. Syst. Res.*, vol. 9, no. 2, pp. 126–163, Jun. 1998.

[36] G. Marakas, R. Johnson, and P. Clay, "The evolving nature of the computer self-efficacy construct: An empirical investigation of measurement construction, validity, reliability and stability over time," *J. Assoc. Inf. Syst.*, vol. 8, no. 1, pp. 16–46, Jan. 2007.

[37] R. Schwarzer and M. Jerusalem, "Measures in health psychology: A user's portfolio," *Causal Control Beliefs*, vol. 1, pp. 35–37, Jan. 1995.

[38] A. Luszczynska, U. Scholz, and R. Schwarzer, "The general self-efficacy scale: Multicultural validation studies," *J. Psychol.*, vol. 139, no. 5, pp. 439–457, Sep. 2005.

[39] S. W. McQuiggan and J. C. Lester, "Diagnosing self-efficacy in intelligent tutoring systems: An empirical study," in *Proc. Int. Conf. Intell. Tutoring Syst.* Berlin, Germany: Springer, 2006, pp. 565–574.

[40] M. El Ouirdi, J. Segers, A. El Ouirdi, and I. Pais, "Predictors of job seekers' self-disclosure on social media," *Comput. Hum. Behav.*, vol. 53, pp. 1–12, Dec. 2015.

[41] Y. Liu, J. Wang, P. Zhao, D. Qin, and Z. Chen, "Research on classification and recognition of driving styles based on feature engineering," *IEEE Access*, vol. 7, pp. 89245–89255, 2019.

[42] A. Kutuzov and E. Kuzmenko, "To lemmatize or not to lemmatize: How word normalisation affects ELMo performance in word sense disambiguation," 2019, *arXiv:1909.03135*.

[43] D. P. Dudău and F. A. Sava, "Performing multilingual analysis with linguistic inquiry and word count 2015 (LIWC2015). An equivalence study of four languages," *Frontiers Psychol.*, vol. 12, p. 2860, Jul. 2021.

[44] N. D. Girsang, "Literature study of convolutional neural network algorithm for batik classification," *Brilliance, Res. Artif. Intell.*, vol. 1, no. 1, pp. 1–7, Sep. 2021.

[45] B. A. M. Ashqar, B. S. Abu-Nasser, and S. S. Abu-Naser, "Plant seedlings classification using deep learning," *Int. J. Academic Inf. Syst. Res.*, vol. 3, no. 1, pp. 7–14, 2019.

[46] A. Farkhod, A. B. Abdusalomov, M. Mukhiddinov, and Y.-I. Cho, "Development of real-time landmark-based emotion recognition CNN for masked faces," *Sensors*, vol. 22, no. 22, p. 8704, Nov. 2022.

[47] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1701–1708.

[48] L. J. Francis and G. Crea, "Happiness matters: Exploring the linkages between personality, personal happiness, and work-related psychological health among priests and sisters in Italy," *Pastoral Psychol.*, vol. 67, no. 1, pp. 17–32, Feb. 2018.

[49] N. Rashid, M. A. F. Hossain, M. Ali, M. I. Sukanya, T. Mahmud, and S. A. Fattah, "AutoCovNet: Unsupervised feature learning using autoencoder and feature merging for detection of COVID-19 from chest X-ray images," *Biocybernetics Biomed. Eng.*, vol. 41, no. 4, pp. 1685–1701, Oct. 2021.

[50] L. He, L. He, and L. Peng, "CFormerFaceNet: Efficient lightweight network merging a CNN and transformer for face recognition," *Appl. Sci.*, vol. 13, no. 11, p. 6506, May 2023.

[51] F. Yuan, L. Zhang, B. Wan, X. Xia, and J. Shi, "Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition," *Mach. Vis. Appl.*, vol. 30, no. 2, pp. 345–358, Mar. 2019.

[52] D. Paul, A. Jain, S. Saha, and J. Mathew, "Multi-objective PSO based online feature selection for multi-label classification," *Knowl.-Based Syst.*, vol. 222, Jun. 2021, Art. no. 106966.

[53] Y. Wang, H. Zhang, and G. Zhang, "CPSO-CNN: An efficient PSO-based algorithm for fine-tuning hyper-parameters of convolutional neural networks," *Swarm Evol. Comput.*, vol. 49, pp. 114–123, Sep. 2019.

[54] P. Singh, S. Chaudhury, and B. K. Panigrahi, "Hybrid MPSO-CNN: Multi-level particle swarm optimized hyperparameters of convolutional neural network," *Swarm Evol. Comput.*, vol. 63, Jun. 2021, Art. no. 100863.

[55] T. A. Judge and J. E. Bono, "Relationship of core self-evaluations traits—Self-esteem, generalized self-efficacy, locus of control, and emotional stability—With job satisfaction and job performance: A meta-analysis," *J. Appl. Psychol.*, vol. 86, no. 1, pp. 80–92, 2001.

[56] T. Zhu, L. Li, J. Yang, S. Zhao, H. Liu, and J. Qian, "Multimodal sentiment analysis with image-text interaction network," *IEEE Trans. Multimedia*, vol. 25, pp. 3375–3385, 2023.

[57] J. Heredia, E. Lopes-Silva, Y. Cardinale, J. Diaz-Amado, I. Dongo, W. Graterol, and A. Aguilera, "Adaptive multimodal emotion detection architecture for social robots," *IEEE Access*, vol. 10, pp. 20727–20744, 2022.

[58] S.-H. Zhang, R. Li, X. Dong, P. Rosin, Z. Cai, X. Han, D. Yang, H. Huang, and S.-M. Hu, "Pose2Seg: Detection free human instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 889–898.

[59] S. P. G. Jasil and V. Ulagamuthalvi, "Deep learning architecture using transfer learning for classification of skin lesions," *J. Ambient Intell. Humanized Comput.*, vols. 1–8, pp. 1–8, Mar. 2021.

[60] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal. Appl.*, vol. 24, no. 3, pp. 1207–1220, Aug. 2021.

[61] G. Akila and R. Gayathri, "Weighted multi-deep feature extraction for hybrid deep convolutional LSTM-based remote sensing image scene classification model," *Geocarto Int.*, vol. 37, no. 27, pp. 18217–18253, Feb. 2024.

[62] H. Gao, S. Peng, and W. Zeng, "Recognition of targets in SAR images using joint classification of deep features fused by multi-canonical correlation analysis," *Remote Sens. Lett.*, vol. 10, no. 9, pp. 883–892, Sep. 2019.

[63] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.

[64] A. Galassi, M. Lippi, and P. Torroni, "Attention in natural language processing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4291–4308, Oct. 2021.

[65] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[66] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, "Multimodal fusion with co-attention networks for fake news detection," in *Proc. Findings Assoc. Comput. Linguistics (ACL-IJCNLP)*, 2021, pp. 2560–2569.

[67] J. L. Tracy and A. C. Weidman, "The self-conscious and social emotions: A personality and social functionalist account," in *Handbook of Personality: Theory and Research*. New York, NY, USA: The Guilford Press, 2021.

[68] Y. Luo, X. Chen, S. Qi, X. You, and X. Huang, "Well-being and anticipation for future positive events: Evidences from an fMRI study," *Frontiers Psychol.*, vol. 8, p. 2199, Jan. 2018.

[69] L. Bujor and M. N. Turliuc, "The personality structure in the emotion regulation of sadness and anger," *Personality Individual Differences*, vol. 162, Aug. 2020, Art. no. 109999.

[70] R. J. Cramer, T. M. S. Neal, and S. L. Brodsky, "Self-efficacy and confidence: Theoretical distinctions and implications for trial consultation," *Consulting Psychol. J., Pract. Res.*, vol. 61, no. 4, pp. 319–334, Dec. 2009.

[71] P. R. Magaletta and J. M. Oliver, "The hope construct, will, and ways: Their relations with self-efficacy, optimism, and general well-being," *J. Clin. Psychol.*, vol. 55, no. 5, pp. 539–551, May 1999.

[72] R. Schwarzer and L. M. Warner, "Perceived self-efficacy and its relationship to resilience," in *Resilience in Children, Adolescents, and Adults: Translating Research Into Practice*. Hoboken, NJ, USA: Wiley, 2013, pp. 139–150.

[73] S. Singh and R. Bala, "Mediating role of self-efficacy on the relationship between conscientiousness and procrastination," *Int. J. Work Organisation Emotion*, vol. 11, no. 1, pp. 41–61, 2020.

[74] A. A. Hayat, K. Shateri, M. Amini, and N. Shokrpour, "Relationships between academic self-efficacy, learning-related emotions, and metacognitive learning strategies with academic performance in medical students: A structural equation model," *BMC Med. Educ.*, vol. 20, no. 1, pp. 1–11, Dec. 2020.

[75] S. Liu, J. You, J. Ying, X. Li, and Q. Shi, "Emotion reactivity, nonsuicidal self-injury, and regulatory emotional self-efficacy: A moderated mediation model of suicide ideation," *J. Affect. Disorders*, vol. 266, pp. 82–89, Apr. 2020.

[76] M. F. Sabri, R. Wijekoon, and H. A. Rahim, "The influence of money attitude, financial practices, self-efficacy and emotion coping on employees' financial well-being," *Manage. Sci. Lett.*, vol. 10, no. 4, pp. 889–900, 2020.

[77] M. Caprara, L. Di Giunta, J. Bermúdez, and G. V. Caprara, "How self-efficacy beliefs in dealing with negative emotions are associated to negative affect and to life satisfaction across gender and age," *PLoS ONE*, vol. 15, no. 11, Nov. 2020, Art. no. e0242326.

[78] D. Motro, D. R. Comer, and J. A. Lenaghan, "Examining the effects of negative performance feedback: The roles of sadness, feedback self-efficacy, and grit," *J. Bus. Psychol.*, vol. 36, no. 3, pp. 367–382, Jun. 2021.

[79] S. Rajput, S. Pande, V. Marda, C. Chandramore, and V. Deokate, "Human activity recognition for surveillance," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2022.

[80] J. Šabić, B. Baranović, and S. Rogošić, "Teachers' self-efficacy for using information and communication technology: The interaction effect of gender and age," *Informat. Educ.*, vol. 21, no. 2, pp. 353–373, Jun. 2021.

[81] Z. Yu, M. Guindani, S. F. Grieco, L. Chen, T. C. Holmes, and X. Xu, "Beyond T test and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in neuroscience research," *Neuron*, vol. 110, no. 1, pp. 21–35, Jan. 2022.

[82] G. Li, Y. Wei, Y. Tian, C. Xu, J.-R. Wen, and D. Hu, "Learning to answer questions in dynamic audio-visual scenarios," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 19086–19096.

[83] R. Arya, A. Kumar, M. Bhushan, and P. Samant, "Big five personality traits prediction using brain signals," *Int. J. Fuzzy Syst. Appl.*, vol. 11, no. 2, pp. 1–10, Apr. 2022.

[84] T. Randhavane, U. Bhattacharya, P. Kabra, K. Kapsaskis, K. Gray, D. Manocha, and A. Bera, "Learning gait emotions using affective and deep features," in *Proc. 15th ACM SIGGRAPH Conf. Motion, Interact. Games*, Nov. 2022, pp. 1–10.

[85] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Sep. 2023.

[86] N. A. Christakis and J. H. Fowler, "Social contagion theory: Examining dynamic social networks and human behavior," *Statist. Med.*, vol. 32, no. 4, pp. 556–577, Feb. 2013.

[87] M. Tschannen-Moran and P. Mcmaster, "Sources of self-efficacy: Four professional development formats and their relationship to self-efficacy and implementation of a new teaching strategy," *Elementary School J.*, vol. 110, no. 2, pp. 228–245, Dec. 2009.

MD. SADDAM HOSSAIN MUKTA received the Ph.D. degree from the Data Science and Engineering Research Laboratory (DataLab), Bangladesh University of Engineering and Technology (BUET), in 2018. He is currently an Associate Professor with the Department of Computer Science and Engineering, United International University (UIU). He has a number of quality publications in both national and international conferences and journals. His research interests include social network analysis and mining, social computing, data mining, and machine learning.

JUBAER AHMAD received the B.Sc. degree in computer science and engineering from United International University (UIU), Dhaka, Bangladesh, in 2022. He is currently a Research Assistant with the IAR Project, UIU. His research interests include computer vision, NLP, big data, and distributed learning.

AKIB ZAMAN received the B.Sc. degree in computer science and engineering from the Military Institute of Science and Technology (MIST), Dhaka, Bangladesh. He is currently pursuing the Ph.D. degree with the Computer Science and Artificial Intelligence Laboratory (CSAIL), Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He has experience in leading research works for premium robotics competitions, such as University Rover Challenge (URC) and U.K.-RAS Medical Robotics Challenge for Contagious Diseases. He is the author of five peer-reviewed publications in international journals and conferences. His research interests include robotics and autonomous systems, reinforcement learning, mining software repositories, image processing, and social computing.

SALEKUL ISLAM (Senior Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Software Engineering, Concordia University, in 2008. He is currently a Professor and the Head of the Department of Computer Science and Engineering, United International University (UIU), Bangladesh. Previously, he was an FQRNT Postdoctoral Fellow with the Énergie, Matériaux et Télécommunications (EMT) Centre, Institut National de la Recherche Scientifique (INRS), Montreal, Canada. His research interests include future internet architecture, blockchain, edge cloud, software-defined networks, multicast security, security protocol validation, machine learning, and AI. He is serving as an Associate Editor for IEEE Access and *Frontiers in High Performance Computing*.