Wiley
IET Biometrics
Volume 2024, Article ID 5351588, 15 pages
https://doi.org/10.1049/2024/5351588

Research Article

Emotion Recognition Based on Handwriting Using Generative Adversarial Networks and Deep Learning

Hengnian Qi,[1] **Gang Zeng**,[1] **Keke Jia**,[1] **Chu Zhang**,[1] **Xiaoping Wu**,[1] **Mengxia Li**,[2] **Qing Lang**,[3] **and Lingxuan Wang**[4]

[1]*School of Information Engineering, Huzhou University, Huzhou 313000, China*
[2]*School of Teacher Education, Huzhou University, Huzhou 313000, China*
[3]*Huzhou University Library, Huzhou University, Huzhou 313000, China*
[4]*Beijing Langcong Technology Co., Ltd., Beijing 100000, China*

Correspondence should be addressed to Qing Lang; 02476@zjhu.edu.cn

Received 23 December 2023; Revised 15 April 2024; Accepted 15 May 2024; Published 27 May 2024

Academic Editor: Ahilan Kanagasundaram

Copyright © 2024 Hengnian Qi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The quality of people's lives is closely related to their emotional state. Positive emotions can boost confidence and help overcome difficulties, while negative emotions can harm both physical and mental health. Research has shown that people's handwriting is associated with their emotions. In this study, audio-visual media were used to induce emotions, and a dot-matrix digital pen was used to collect neutral text data written by participants in three emotional states: calm, happy, and sad. To address the challenge of limited samples, a novel conditional table generative adversarial network called conditional tabular-generative adversarial network (CTAB-GAN) was used to increase the number of task samples, and the recognition accuracy of task samples improved by 4.18%. The TabNet (a neural network designed for tabular data) with SimAM (a simple, parameter-free attention module) was employed and compared with the original TabNet and traditional machine learning models; the incorporation of the SimAm attention mechanism led to a 1.35% improvement in classification accuracy. Experimental results revealed significant differences between negative (sad) and nonnegative (calm and happy) emotions, with a recognition accuracy of 80.67%. Overall, this study demonstrated the feasibility of emotion recognition based on handwriting with the assistance of CTAB-GAN and SimAm-TabNet. It provides guidance for further research on emotion recognition or other handwriting-based applications.

1. Introduction

Among negative emotions, sadness is a pervasive and common everyday emotion. Negative emotions can drain individuals of their talents, damage their physical and mental health, and hinder their careers. Prolonged negative emotional states can lead to excessive mental stress, reduced productivity, compromised immunity, and even suicide [1, 2]. Several researchers have focused on identifying negative emotions and studying their impact on mental health [3, 4, 5, 6, 7].

Handwriting, as a unique trace left by the writer during the writing process, reflects the interplay between subjective factors and objective conditions. Handwriting is a natural and authentic mark that retains individuality, unaffected by artificial influences [8, 9]. It is a reflex activity of the brain in response to external stimuli related to our internal conditions, including emotional and psychological factors [10, 11, 12].

Handwriting analysis, also known as graphology, is a scientific method of determining, assessing, and understanding personality traits based on the strokes and patterns revealed by handwriting. It can provide insights into the writer's mental state, including emotional expressions, anxieties, and honesty, among other feelings [13]. Handwriting is frequently utilized to determine various traits of a person, such as personality, neurodegenerative diseases, emotional states, gender, age, nationality, and more [14, 15, 16, 17, 18, 19, 20]. Handwriting is a distinctive feature of each individual, offering as much uniqueness as fingerprints [21]. For instance, an analyzed handwriting sample may indicate a high possibility that the author was under stress during writing, which can be helpful in investigations. While the suggested methodologies

may not discriminate between different types of stress (e.g., self-imposed or explicit threats), they can provide quantifiable and reproducible indicators for the level of stress a user is experiencing, thereby providing crucial, though not always absolute, evidence [22]. Researchers like Sen and Jing-Xiao [10] have analyzed handwriting samples collected from individuals in various emotional states, such as happiness, anger, sadness, fear, and calmness, and discovered significant differences and patterns in line spacing, writing strength, font size, and typos. Understanding the association between handwriting and emotions assists in comprehending the emotional traces left by writers. Other researchers, such as Nolazco-Flores et al. [23], have combined writing and drawing features to identify anxiety and tension. Additionally, Kedar et al. [24] have used handwriting features such as baseline, slant, size, pen pressure, zone, and margin to determine a person's emotional level, aiding in the diagnosis of those who require psychological assistance due to mental distress or depression. Furthermore, Cordasco et al. [25] have studied the use of handwriting and drawing characteristics to detect negative emotions. They collected traits from individuals experiencing depression, stress, and anxiety and assessed them using the DASS-42 scale, capturing handwriting and drawing samples with a digital tablet.

Emotion elicitation is a research technique used to induce typical emotions in a laboratory setting through specific procedures such as viewing emotionally charged film clips, pictures, or listening to emotionally charged music. Combining audio-visual stimuli, such as film clips or music videos, can more effectively evoke corresponding emotions in subjects [26, 27, 28, 29]. The Self-Assessment Manikin (SAM) is commonly used to measure emotional arousal and assess subjects' self-rated emotional states [30, 31]. Researchers like Fairhurst et al. [32] have employed SAM to measure emotional states after elicitation.

Recent advancements in deep learning have opened up new avenues for emotion recognition, offering improved feature extraction, selection, and classification capabilities compared to conventional methods. However, collecting sufficient handwriting datasets within a limited budget can be challenging. Generative adversarial networks (GANs) have shown promise in data synthesis and have been applied to generating tabular data [33, 34, 35]. GAN can enhance sparse datasets with low sample sizes and poor quality, while also providing privacy guarantees. Additionally, TabNet, a neural network structure, combines tree models with deep neural networks and has been proposed for tabular data classification [36]. The attention mechanism module SimAM allows for better flexibility and effectiveness in capturing salient features [37].

Based on the above background, this study aims to explore emotion recognition using dynamic information from handwriting. Negative emotion (sadness) and nonnegative emotions (happiness and calmness) are selected for analysis. Emotion data is collected through an emotion elicitation procedure, followed by the synthesis of sample data using GANs. Finally, machine learning and deep learning methods (specifically TabNet with SimAM) are applied to analyze the synthesized sample data for emotion recognition.

2. Materials and Methods

2.1. Data Collection Experimental Setup

2.1.1. Experimental Equipment. The equipment used in this experiment includes a dot-matrix pen and dot-matrix paper (Tstudy China). The dot-matrix paper consists of several tiny dots arranged according to a specific algorithmic pattern (i.e., Anote's patented dot-matrix pattern printed on regular paper). These dots provide the digital pen with coordinate parameters, ensuring accurate recording of handwriting during writing. The dot-matrix digital pen is composed of a pressure sensor, high-speed camera, memory, processor, battery, and Bluetooth or USB communication module. When the pen tip is pressed, the pressure sensors are triggered, activating the built-in high-speed camera, which captures images of the dot-matrix at a rate of 100 frames per second as the pen tip moves. The captured information, including dot-matrix coordinates, stroke order, pressure data, and movement speed, is transmitted to the built-in processor, which eventually outputs the data via Bluetooth or USB communication. The dot-matrix pen and paper used in this experiment are shown in Figure 1.

2.1.2. Subject Population. Forty-four students from Huzhou University were randomly invited as subjects for this experiment through recruitment. The 44 subjects were made up of 23 males and 21 females. These students were between 22 and 26 years old. They all had a bachelor's degree or above. They were healthy, with no physical or psychological illnesses. They were right-handed, had normal vision or corrected vision, and had normal hearing.

2.1.3. Experimental Content and Procedures. The text material chosen for this experiment is a neutral text passage with no particular meaning and a smooth flow of words. Each paragraph is about 60 words, comprising two sections of text. The text was written as follows:

Text 1: "A color screen has been added to the Xiaomi bracelet for the naked eye. The display is detailed enough that you cannot even see the particles when you lift your hand to scan it. Under the color screen, some animations give the whole bracelet a premium feel."

Text 2: "The color screen also brings another problem that is hard to ignore, and that is the screen brightness; the Xiaomi bracelet has five levels of brightness settings, indoors probably at three levels will be more comfortable; at night, you can automatically turn on the night mode to solve."

Each subject completes three experimental tasks on the same day: "Natural State" (Task 1), "Inducing Happy" (Task 2), and "Inducing Sad" (Task 3). When the subjects arrived in the laboratory, they were first registered with their name, gender, and age. Then, the research background is explained, and the experimental equipment is distributed.

In Task 1, the subject was first asked to sit still for 2 min and then put in a natural state. The subject was asked to copy Text 1 and Text 2 with a dotted pen and rate the current emotion on a scale of 0–10 on two dimensions: happy and sad, with 0 indicating no feeling and 10 indicating the most intense emotional experience, Task 1 was all completed with the subject in a calm mood.

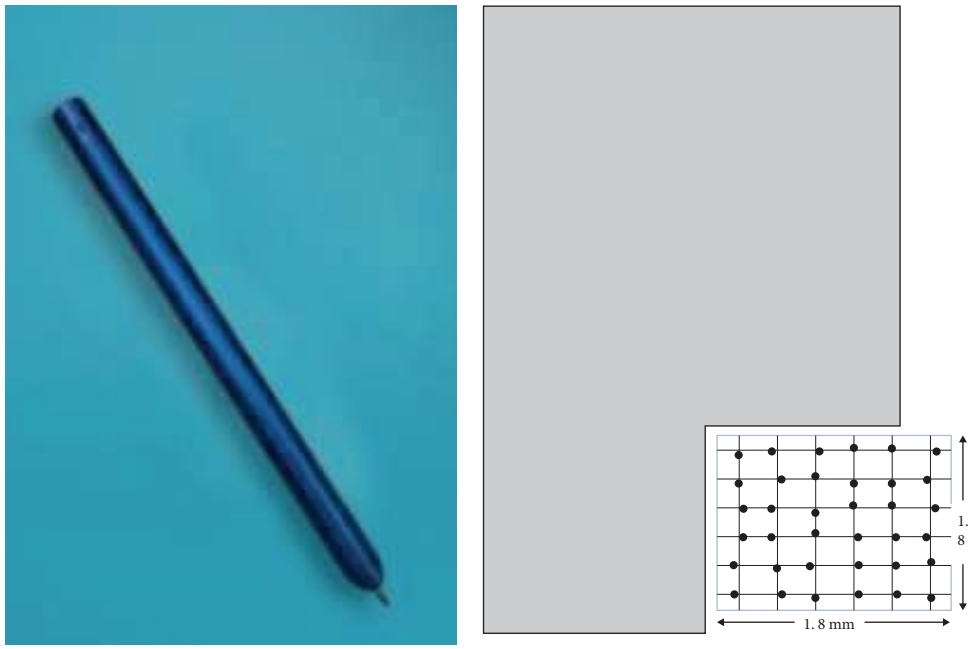

Figure 1: Dot-matrix equipment: (a) dot-matrix pen diagram; (b) dot-matrix paper diagram.

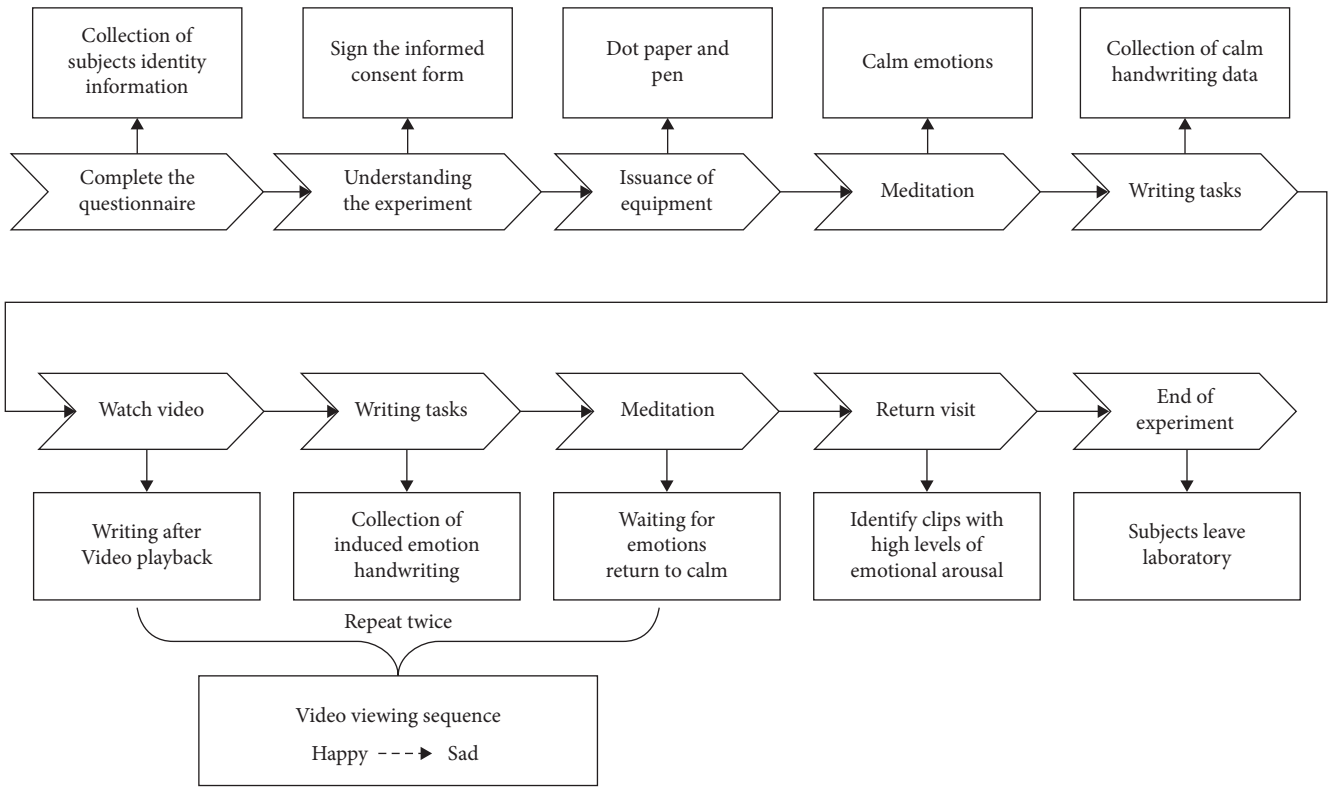

Figure 2: Specific details of the process for subjects to participate in the data collection tasks.

In the "Inducing Happy" (Task 2) and "Inducing Sad" (Task 3) experiments, subjects watched a video material of ~9–10 min in length and transcribed Text 1 and Text 2 with a dotted pen at the end of the video. The emotions evoked by the video clip were rated on a scale of 0–10. The flow of the experiment is shown in Figure 2. In Task 2, the majority of subjects (29) scored "happy"; in Task 3, the majority of subjects (28) scored "sad"; the "happy" and "sad" scores in Task 1

were lower than those in Task 2 and Task 3, and the result was a "Calm" emotional state.

2.2. Feature Extraction and Preprocessing. This experiment was conducted by sampling the 44 subjects mentioned above and having them write with dot-matrix paper and pen to obtain digital handwriting. A total of 48 features were obtained, all of which were brought by the lattice pen's instantaneous pressure, the point's coordinate value, and the time stamp. Stroke characteristics correspond to the intermediate state of features when a stroke is completed. The task characteristics correspond to the intermediate state of the features at the time of writing completed the task. Among them, the degree of continuous stroke only appears in the task characteristics, and the other 47 features are involved in the stroke and task. Then, a list of handwriting features was extracted, as shown in Table 1.

The degree of consecutive strokes (L_{pen}) is the ratio of the number of missing strokes to the number of strokes in the task, which can be expressed as in Equation (1): the number of strokes written in the whole writing task is the number of actual strokes written by the subject (W_{num}). The number of missing strokes is the difference between the total number of strokes in the task (N_{stroke}) minus the number of actual strokes written.

$$L_{\text{pen}} = \frac{N_{\text{stroke}} - W_{\text{num}}}{N_{\text{stroke}}}. \tag{1}$$

The average pressure is expressed as in Equation (2): the writing pressure is the pressure-sensitive data obtained from the 1,024-level pressure sensor built into the dot-matrix pen; a higher pressure-sensitive value means a higher writing pressure.

$$\overline{f} = \frac{1}{n}\sum_{i=1}^{n} f_i. \tag{2}$$

The entropy value is expressed as in Equation (3): information entropy commonly characterizes the information's uncertainty. The lower the entropy value, the higher the signal self-similarity and the lower the complexity. ($P(x_i)$) stands for the probability that the random variable x takes the value x_i. It is a probability distribution function indicating the likelihood of each possible value occurring.

$$E(x) = -\sum_{i=1}^{n} P(x_i)\log_2(P(x_i)). \tag{3}$$

The lag time for one task is as in Equation (4): (t_i) is the time between the end of one stroke and the start of the following stroke, and the lag time (T_{lag}) for this task is the sum of t_i.

$$T_{\text{lag}} = \sum_{i=1}^{n} t_i. \tag{4}$$

The variance is expressed as in Equation (5): a measure of the degree of dispersion when reflecting a set of data, i.e., the degree of deviation (M) is the average.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2. \tag{5}$$

The average speed ($\overline{V_w}$) is expressed as in Equation (8): reflecting the writing rate, the average speed for the whole task. As shown in Equation (6), (L_w) is the sum of the lengths of all strokes in the whole task. As shown in Equation (7), (T_w) is the time it takes to complete all strokes in the whole task.

$$L_w = \sum_{i=1}^{n} l_i, \tag{6}$$

$$T_w = \sum_{i=1}^{n} t_i, \tag{7}$$

$$\overline{V_w} = \frac{L_w}{T_w}. \tag{8}$$

Since handwritten data from different subjects are heterogeneous, they need to be transformed into a common domain for further analysis. The preprocessing method used in this experiment is z-score, and the z-score formula is expressed as in Equation (9). Where (X^*) is the mean and variance normalized sample of characteristics, (x) is the sample of parts, (μ) is the mean of the characteristic data, and (σ) is the standard deviation of the typical data.

$$X^* = \frac{x - \mu}{\sigma}. \tag{9}$$

2.3. Model Construction. In total, 210 (there were 35 valid samples, 3 tasks, and 2 segments of text among the 44 subjects) task-based samples and 60,000 stroke-based samples were obtained. Among the task-based samples, the number of "calm" samples was 82, the number of happy pieces was 74, and the number of low samples was 54; the number of "calm" samples was about 22,000, the number of comfortable models was ~21,000, and the number of familiar pieces was about 19,000. The total number of fragments in the task-based sample was small, and the sample size was expanded by generating an adversarial network (conditional tabular-GAN (CTAB-GAN)). A deep learning model based on the attention mechanism was then constructed for emotion classification.

2.3.1. Generating Adversarial Network Model Construction. The GAN model selected in this paper is CTAB-GAN, CTAB-GAN (a novel conditional table GAN), which is proposed as an efficient way of synthesizing tabular data [34].

Table 1: Detailed features extracted from handwriting.

Number	Features
1	The degree of consecutive strokes
2	Average pressure
3	Pressure entropy value
4	Lag time
5	Variance of pressure
6	The standard deviation of pressure
7	Stroke time
8	Stroke length in x-direction
9	Stroke length in y-direction
10	Stroke length
11	x-Coordinate variance
12	y-Coordinate variance
13	The standard deviation of x-coordinate
14	y-Coordinate standard deviation
15	Average velocity
16	x-Axis mean velocity
17	The average velocity on the y-axis
18	Stroke velocity maximum
19	Stroke velocity minimum
20	Entropy of velocity
21	Number of times the rate has slowed down
22	Variance of velocity
23	The standard deviation of velocity
24	The maximum value of speed in the x-axis direction
25	The minimum value of speed in the x-axis direction
26	The maximum velocity in the y-axis
27	The minimum value of momentum in the y-direction
28	The variance of rate in the x-axis direction
29	The standard deviation of rate in the x-axis
30	y-Axis velocity variance
31	The standard deviation of velocity in the y-axis direction
32	Average acceleration
33	Average acceleration on the x-axis
34	Average acceleration on the y-axis
35	Acceleration maximum
36	The minimum value of acceleration
37	Entropy of acceleration
38	The number of times acceleration has slowed down
39	Variance of acceleration
40	The standard deviation of acceleration
41	The maximum value of acceleration in the x-axis direction
42	The minimum value of acceleration in the x-axis direction
43	The maximum weight of acceleration in the y-axis direction
44	The minimum value of acceleration in the y-direction
45	The variance of acceleration in the x-axis direction
46	The standard deviation of acceleration in the x-axis direction
47	y-Axis acceleration variance
48	The standard deviation of acceleration in the y-axis direction

The GAN architecture consists of three modules: the generator, the discriminator, and the auxiliary classifier. A set of noise vectors are added to the generator to form synthetic data. The synthetic data and the original data are judged in the discriminator whether the data is the data in the training set, the true data is labeled as 1, and the false data is labeled as 0. The training stops when the discriminator cannot distinguish the synthetic data from the original data. Real data and synthetic data are subjected to additional supervision through a classifier. To improve performance in ML applications. Continuous, categorical, and mixed variables are efficiently modeled through novel data encoding and conditional vectors, which are better suited for task-based sample expansion work. The encoder adopts an innovative approach by systematically encoding mixed-type variables. For the continuous part, we employ a variational Gaussian mixture (VGM) model [38], treating the values of mixed variables as concatenated value-mode pairs. This approach effectively handles mixed variables, allowing them to capture categorical characteristics while retaining the distributional information of continuous values. Specifically, a VGM model is utilized to estimate the number of modes in the continuous part and fit a Gaussian mixture. To encode values in the continuous part, we select the mode with the highest probability and standardize the values. Simultaneously, through one-hot encoding, we retain the mode information used for encoding, enhancing the interpretability of the model. For categorical variables, a similar encoding strategy is employed, representing mode values directly through one-hot encoding. Additionally, special consideration is given to missing values, treating them as a distinct and independent category. The final encoding is obtained by concatenating the encoded continuous and categorical parts, providing the model with comprehensive modeling capabilities for mixed-type data. This novel data encoding method furnishes the sentiment recognition model with clearer and more effective inputs, particularly well-suited for handling real-world scenarios involving mixed-type variables. The process and detailed information of CTAB-GAN synthesis based on original emotional data is illustrated in Figure 3. The generator and discriminator are implemented as four and two-layer CNNs, respectively, leveraging CNNs' ability to capture relations between pixels in an image to enhance semantic integrity [39]. The classifier, with seven layers, is trained on original data to interpret semantic integrity better. Synthetic data undergo a reverse transformation from their encoding before being input into a classifier for class label predictions.

2.3.2. Deep Learning Model with Attentional Mechanisms. The deep learning model chosen for this experiment was TabNet, a neural network specifically for tabular data that uses the idea of sequential attention to mimic the behavior of a decision tree [36]. It can be seen as a multistep neural network, using two critical operations at each step: feature transformer and attentive transformer. The attention mechanism is incorporated as SimAm (a simple attention module without parameters), placed between the feature transformer and attentive transformer in TabNet, facilitating the acquisition of an optimized weight matrix surpassing the target. The

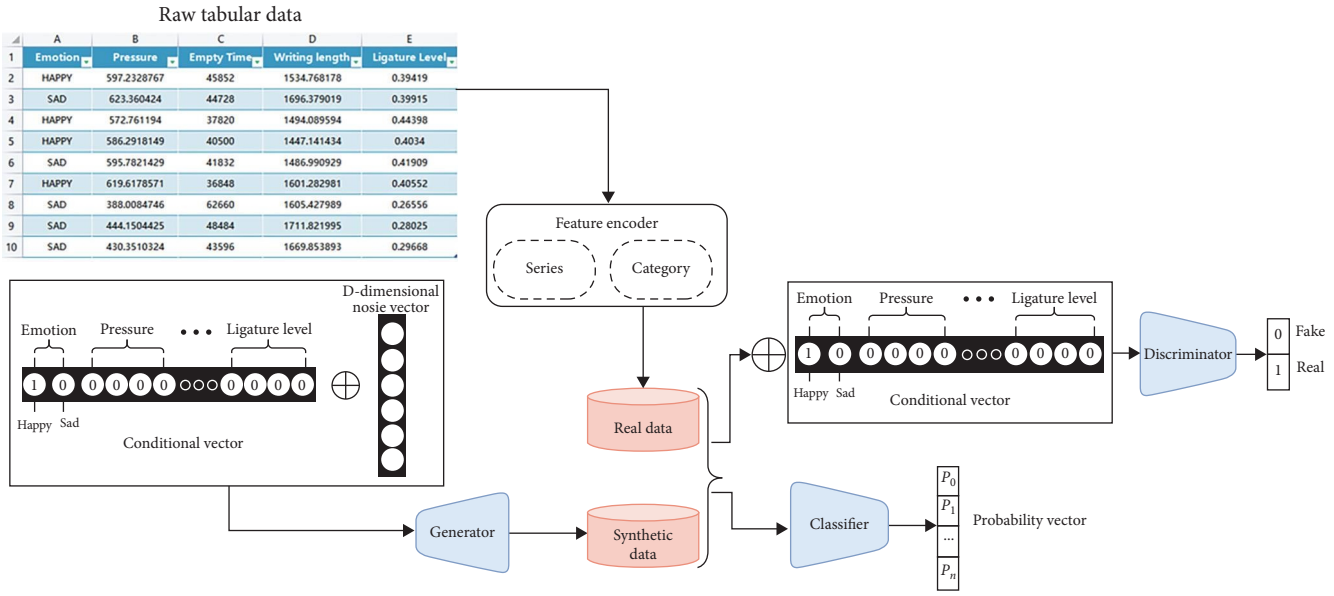

FIGURE 3: Synthesis of handwriting-based mood data via CTAB-GAN.

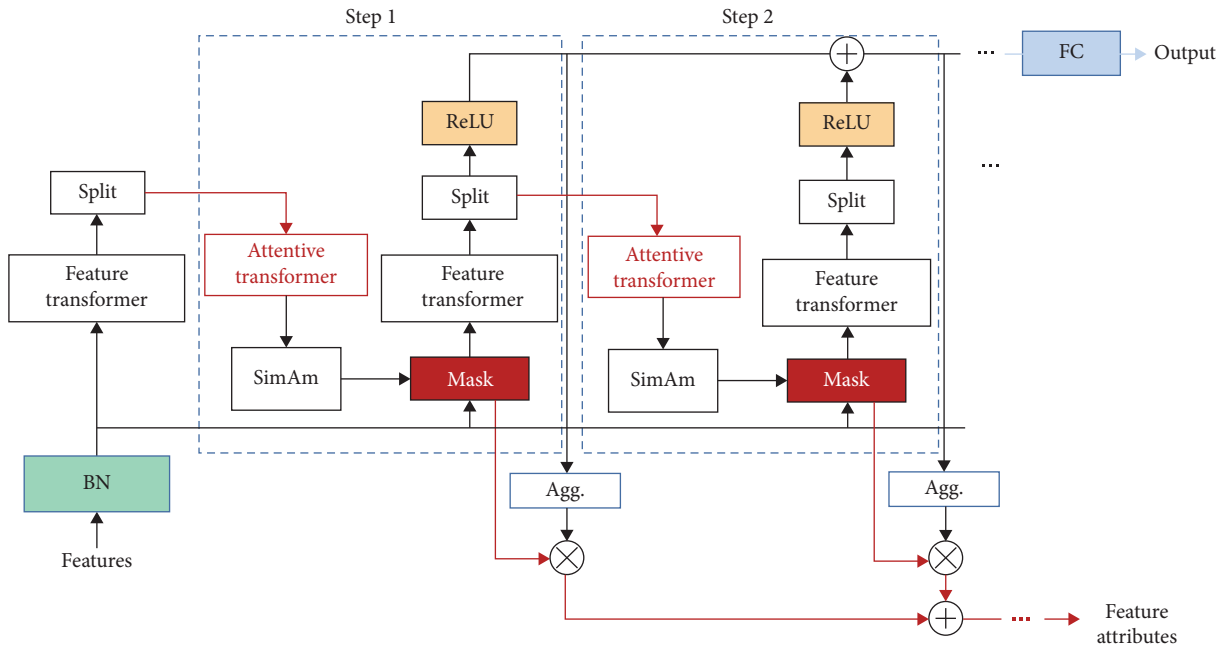

FIGURE 4: Flowchart of TabNet incorporating the attention mechanism SimAm.

feature data enters the BN (batch normalization) layer for data normalization. The result is used for feature calculation and enters the feature selection part for weight assignment. The essential features are filtered through the weight matrix to enter segment the analysis again, followed by split segmentation processing, and part of it is used as common standards after the activation function. After all the steps are completed, the common features are used to complete the final decision. The other part goes to the next step and learns the personality traits of each step, and so on until all decision steps have been completed. The model can achieve end-to-end learning while selecting and processing the most valuable features, thus improving interpretability and learning ability, and the SimAm-TabNet flowchart is shown in Figure 4.

Feature transformer: To maintain parameter effectiveness and robust learning, the feature transformer consists of layers shared between decision steps and layers dedicated to the current decision step. This design reduces parameter quantity compared to scenarios where all parameters only act on the current decision step, enhancing robustness. In contrast to designs where every shared layer has identical capabilities in each decision step, this approach allows

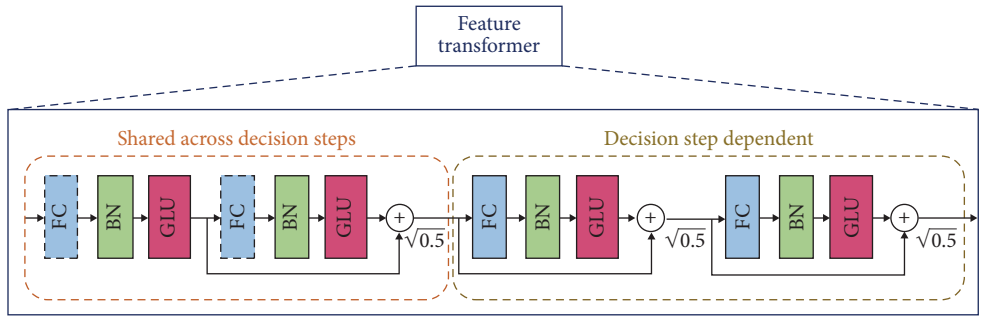

Figure 5: A feature transformer block example—4-layer network is shown, where 2 are shared across all decision steps and 2 are decision step-dependent. Each layer is composed of a fully connected (FC) layer, BN, and GLU nonlinearity.

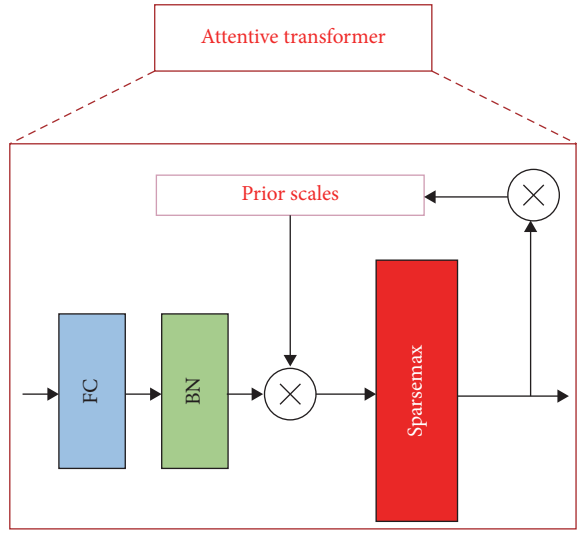

Figure 6: An attentive transformer block example—a single-layer mapping is modulated with a prior scale information, which aggregates how much each feature has been used before the current decision step. Sparsemax is used for the normalization of the coefficients, resulting in the sparse selection of the salient features.

parameters to possess varied feature processing capabilities in each decision step, contributing to more effective feature processing at each step. To stabilize the training process and prevent drastic variance changes, each residual link is multiplied by $\sqrt{0.5}$. The activation function employed is the gated linear unit (GLU) gating mechanism activation function. GLU is a commonly used nonlinear activation function in neural networks. The schematic diagram of the feature transformer is illustrated in Figure 5.

Attentive transformer: Attentive transformer achieves the selection of features for the current decision step by learning a weight matrix. Initially, the schematic diagram of the attentive transformer is illustrated in Figure 6. Data (a [i–1]) passes through a fully connected layer and batch normalization layer (h_i). The output of the h-layer is multiplied by the prior scale of the previous decision step (P [i−1]) and then undergoes the sparsemax activation function to generate the required M [i]. This process completes the feature selection. The formula for learning the weight matrix is as follows:

$$M[i] = \text{sparsemax}\left(P[i-1] \cdot h_i\left(a[i-1]\right)\right). \tag{10}$$

The SimAm was selected as an attention mechanism to be used in the TabNet, which is a nonreferential attention mechanism module [37]. To better implement attention, the importance of each neuron needs to be evaluated, and an energy function is defined to minimize the energy function. The more minor the energy, the more significant the difference between the target neuron and other neurons, and the higher the importance. The minimum energy formula is shown in Equation (11). (e_t^*) is a computed energy value associated with a target neuron, indicating the difference between the target neuron and other neurons. A lower value signifies higher importance. ($\widehat{\sigma}^2$) is an estimated variance. (λ) is a parameter used in the energy function. (t) represents the target neuron. ($\widehat{\mu}$) is an estimated mean.

$$e_t^* = \frac{4(\widehat{\sigma}^2 + \lambda)}{(t - \widehat{\mu})^2 + 2\widehat{\sigma}^2 + 2\lambda}. \tag{11}$$

The importance of the neuron is ($1/e^*$), and after obtaining the significance, the feature matrix is augmented with the formula shown in Equation (12). (E) is the significance of a

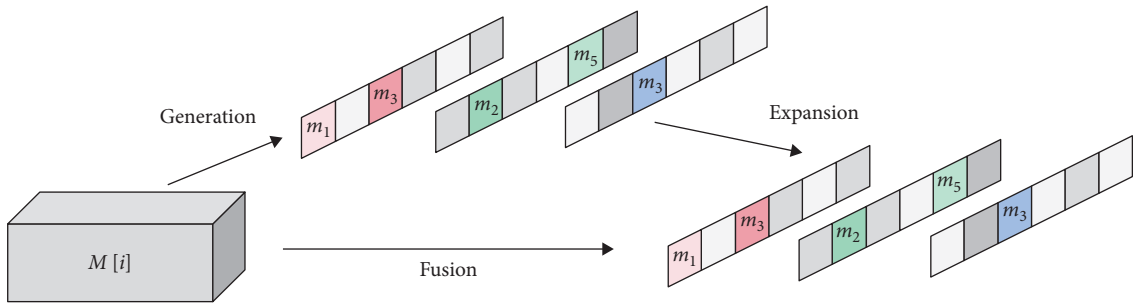

Figure 7: Schematic representation of the feature matrix obtained by the attention mechanism SimAm.

neuron, calculated as the reciprocal of e_t^*. (X) is The feature matrix. $(\tilde{X})$ is the augmented feature matrix after applying the sigmoid activation function and element-wise multiplication with the significance values.

$$\widetilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X. \tag{12}$$

This study used SimAm-TabNet based on the original TabNet with the addition of the attention mechanism SimAm between feature selection and feature calculation so that TabNet gets a better weight matrix than the target from the weights of the features. SimAm gets the feature matrix in TabNet based on the importance of neurons, as shown in Figure 7.

3. Results

3.1. Analysis of Data Collection Results. A total of 44 subjects participated in this study, and handwriting data was collected from the subjects through emotional elicitation. Among them, 23 male students were in the age range of 21–27 years, and 21 female students were in the age range of 20–26 years.

There were 44 self-rated scales for 44 subjects, and the scores on the self-rated scales were the scores on the current mood dimensions. The results of the subjects' mood scale scores are shown in Table 2.

The subjects were all rated on an emotion scale after the emotion induction, ranging from 0 to 10. A score of 0 indicated no emotion arose, and 10 indicated the most intense emotional experience. As seen from Table 2, the majority of subjects in both Task 2 and Task 3 scored results greater than 7, demonstrating that the stimulus material selected was indeed effective in evoking the desired emotional state. However, there were still inevitably a few subjects whose emotions were more difficult to evoke and scored lower. Therefore, the corresponding handwriting data were not suitable for use in emotion recognition studies, and these data were discarded.

In this study, data with scores of 0–6 on both the "happiness" and "sadness" dimensions in Task 1 were regarded as "calm." Data with scores of 0–6 on both Task 2 and Task 3 were regarded as invalid and discarded. Only data with scores of 7–10 were considered valid and used for subsequent classification studies. The emotion scale scores of the correct data were counted separately, and a total of 35 effective samples were obtained, and the statistical results are shown in Table 3.

As seen in Table 3, the emotionally evoked scores of Task 2 and Task 3 were higher than those of Task 1, and the standard deviation values were smaller. This indicates that the overall effect of the emotion elicitation experiment was good and that the degree of emotion evocation was high. Therefore, the handwriting data collected does contain the required emotion-related information and can be used in the follow-up study.

3.2. GAN Synthesis Data Results. In total, 210 pieces of original data (82 for calm, 74 for happy, and 54 for sad) were synthesized into 4,444 bits of data (1,487 for calm, 1,477 for happy, and 1,480 for sad) by CTAB-GAN. A comparison was made between original and synthetic data using two widely adopted machine learning models and machine learning model evaluation metrics, including accuracy, balanced *F*-score, and AUC values. Both machine learnings used default parameters in scikit-learn 0.23.2. The ML evaluation process is illustrated in Figure 8.

The "Happy" and "Sad" data were compared with the original and synthetic data in terms of Accuracy, balanced *F*1-Score, and AUC value evaluation metrics in Random Forest and SVM, as shown in Table 4. Sample trends for the original and synthetic data for "Happy" are shown in Figure 9. The assessment results in Table 4 and Figure 9 show that the difference in assessment metrics between the synthetic and real data is slight, and the data trends are essentially the same. Synthetic data can be used in the study of emotion recognition.

3.3. Emotional State Prediction Results. This paper used tasks to explore different emotional states. In the following experiment, the sample size of original data with task unit was 210 (82 for calm, 74 for happy, and 54 for sad), and CTAB-GAN generated the composite data. The synthesized task sample data volume was 4,444 (1,487 for calm, 1,477 for happy, and 1,480 for sad). The specific experiment is as follows.

Task-based samples were used to experimentally explore negative and nonnegative emotions. Among them, sad emotion is regarded as a negative emotion; 30 and 35 samples from calm and happy emotions are selected as nonnegative emotions. In the original data, there were 54 negative samples and 65 nonnegative samples. In the synthetic data, there were 1,480 negative samples and 1,477 nonnegative samples.

Table 2: Results of emotion scale scores.

No.	Types of emotional triggers			
	Happy		Sad	
	Task 1	Task 2	Task 1	Task 3
1	6	9	6	8
2	9	10	1	8
3	2	7	1	7
4	6	8	1	9
5	6	4	5	7
6	3	9	2	9
7	2	9	2	6
8	4	7	2	8
9	10	10	1	2
10	3	8	2	9
11	6	6	1	4
12	7	9	2	3
13	3	8	2	8
14	9	10	1	8
15	3	8	1	10
16	6	8	2	5
17	2	3	2	4
18	6	6	5	9
19	2	8	1	7
20	2	9	3	7
21	8	9	2	10
22	10	10	1	5
23	3	4	1	4
24	4	7	8	7
25	6	8	3	6
26	1	2	1	2
27	2	8	1	9
28	4	7	2	7
29	1	4	1	6
30	6	6	1	6
31	2	3	5	6
32	6	8	4	6
33	3	8	2	9
34	9	10	2	6
35	5	7	5	9
36	4	5	2	8
37	2	5	1	8
38	3	8	2	7
39	3	6	1	8
40	1	2	1	2
41	3	6	2	5
42	2	5	1	9
43	6	8	2	7
44	2	9	2	9
Mean	4.386	7.068	2.182	6.795
SD	2.563	2.214	1.603	2.141

The original and synthetic data on negative emotions and nonnegative emotions were compared to machine learning and deep learning models. The ratio of the training set to the test set is 7 : 3. In our experiments, we employed 10-fold cross-validation, ensuring an adequate number of training iterations for each model on every subset to mitigate variations arising from randomness. The training and validation steps for each model were repeated 10 times to attain more stable and reliable performance evaluations. Regarding the parameter configurations for SimAm-TabNet, the settings are as follows:

Table 3: Statistical results of effective scoring.

	Types of emotional triggers			
	Happy		Sad	
	Task 1	Task 2	Task 1	Task 3
Mean	4.343	7.857	2.314	7.429
SD	2.274	1.556	1.787	1.659

Input layer neurons: The number of neurons in the input layer, equivalent to the number of features, is set to 48 in this study, representing the raw input features.

Output layer neurons: For the binary classification task, it is set to 2.

Decision steps: This parameter determines the size of the network structure steps and is set to 8 in this paper.

Predictive phase feature number: It corresponds to the number of features inputted in each decision step. In this study, it is set to 30.

Feature selection phase feature number: It corresponds to the number of features outputted in each decision step, set to 10 in this paper.

Batch sample processing (batch size): It is set to 200.

Optimizer: The optimizer is set to Adam.

Learning rate: The learning rate for SimAm-TabNet is set to 0.001.

Epochs: It is set to 500.

The loss function: cross-entropy.

The accuracy results of the training set and test set are shown in Table 5.

The usability of the generated adversarial network data was tested under the machine learning and SimAm-TabNet models by using the synthetic samples (1,480 negative samples and 1,477 nonnegative samples) and the original samples (54 negative samples and 65 nonnegative samples). The confusion matrices for the synthetic samples and the original samples are shown in Figure 10. By comparing the confusion matrices of the training and testing sets, it can be observed that the accuracy between the two sets is similar, indicating that SimAm-TabNet did not exhibit overfitting. Additionally, the presence of overfitting can also be inferred from the behavior of the loss function. The loss curves of SimAm-TabNet are shown in Figure 11; with an increase in the number of training iterations, the classification performance of both the training and testing sets stabilizes and eventually converges to similar values. Combining this observation with the confusion matrices of SimAm-TabNet

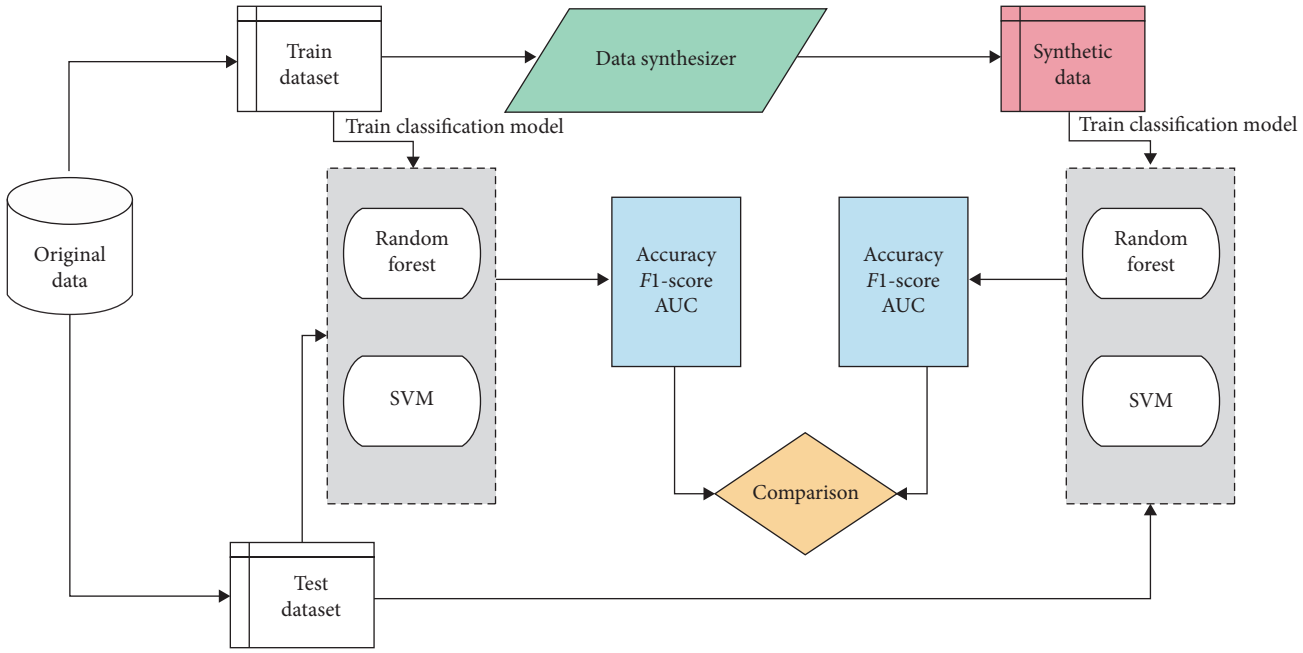

Figure 8: Evaluation process of original data and synthetic data in ML.

Table 4: Evaluation results of original and synthetic data.

	Accuracy	*F*1-score	AUC
Real			
SVM	0.7672	0.7593	0.8024
Random forest	0.7912	0.7855	0.8284
Fake			
SVM	0.7821	0.7754	0.8456
Random forest	0.8204	0.8122	0.8634

further confirms that the model did not experience overfitting. The classification results of the synthetic samples and the original samples are shown in Table 6.

Cohen's Kappa coefficient is a statistical measure used to assess the level of agreement between two raters or classifiers, taking into account the possibility of agreement occurring by chance. It is particularly useful when evaluating classification performance, providing a more robust metric than simple accuracy, especially in situations with imbalanced class distributions. The formula for calculating Cohen's Kappa is as follows: P_o represents the observed agreement, and P_e signifies the expected agreement.

$$K = \frac{P_o - P_e}{1 - P_e}. \tag{13}$$

As can be seen through Tables 5 and 6, for negative emotion detection, SimAm-TabNet outperformed the traditional machine learning models, with the classification results of 80.67%, a Kappa value of 0.62 can be considered a relatively good level of consistency. Synthetic data from the GAN was used to train the model, which also tested the real dataset with good negative emotion detection.

4. Discussion

By analyzing the results of emotion recognition using task sample features, it can be observed that task sample features effectively help us identify the emotional state of the writer. This is because the task unit better reflects the overall stable characteristics of the writer, allowing the writer's handwriting features to more consistently reflect the changes produced by internal conditions.

The use of GANs improved the recognition accuracy of task samples by 4.18%. GAN was used to augment the sample set of tabular data, resulting in the SimAm-TabNet model outperforming machine learning models in terms of metrics. GAN proves to be a successful solution for addressing the issue of insufficient data for model training, and it helps us generate data sets similar to the original data samples for training models to predict emotional states. In situations where collecting real samples is challenging, GAN can assist in increasing the number of samples that match the experimental data. By introducing the attention mechanism, the

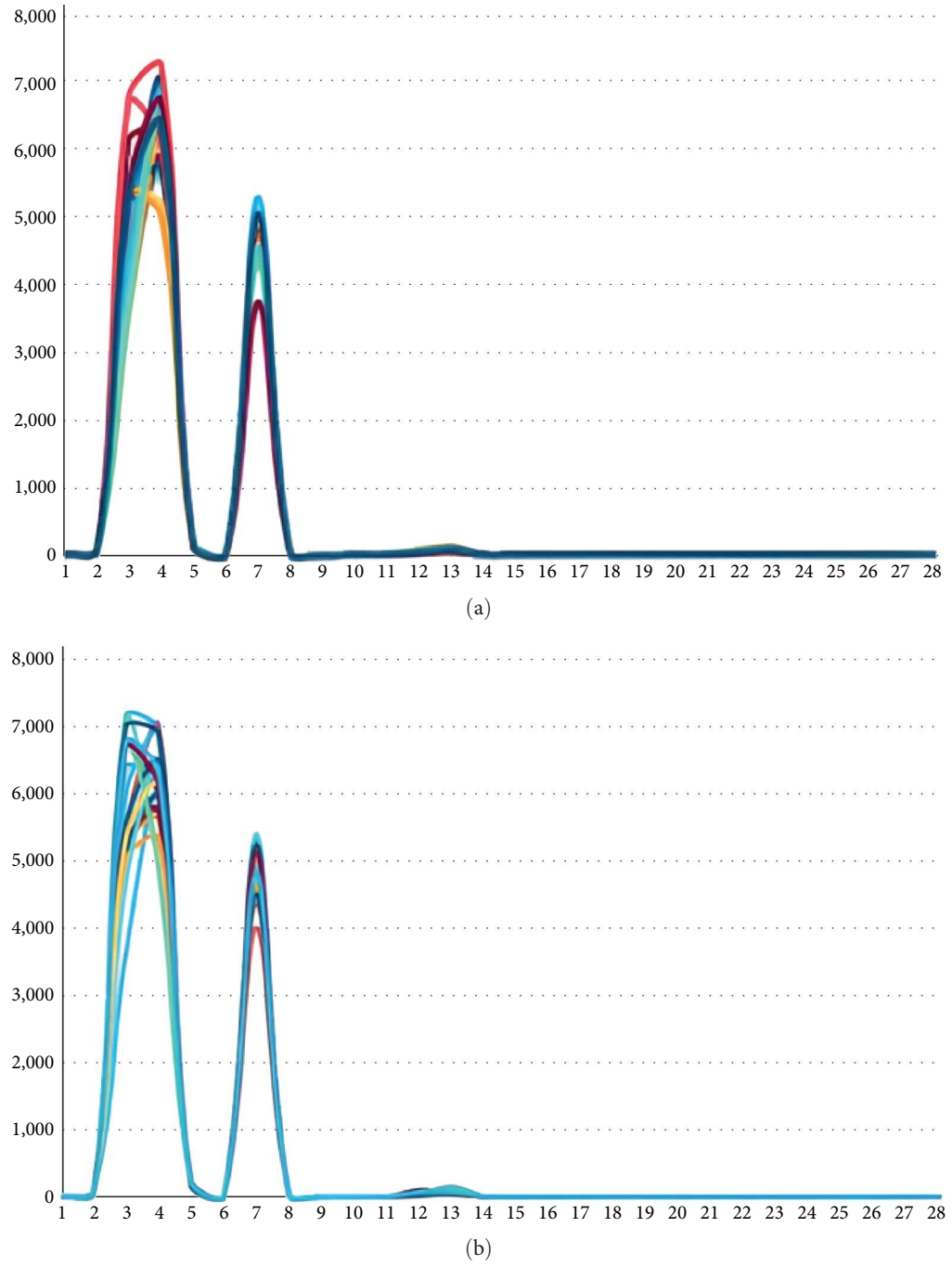

Figure 9: "Happy" original and synthetic data trend graphs: (a) original data; (b) synthetic data.

original TabNet deep learning model can focus more on key features; the incorporation of the SimAm attention mechanism led to a 1.35% improvement in classification accuracy. In comparison to other methods, such as the popular CNNs, LSTM, and optimized LSTM deep learning algorithms [40, 41], the TabNet model demonstrates superior performance in handling tabular data. Tabular data may lack clear semantic information due to context, yet the TabNet algorithm excels in end-to-end learning while automatically selecting and processing the most relevant features. Particularly noteworthy is our extraction of 48 handwriting features, far surpassing the data available in the public EMOTION database. Additionally, with the incorporation of the SimAm attention mechanism, the model can better select the most useful features from the tabular data, further enhancing the performance and generalization ability of the model.

The preliminary results of using Chinese handwriting data to predict emotional states show feasibility. This is because, in the emotional space model, "sadness" is positioned on the negative valence axis, representing negative emotions, while "happiness" is positioned on the positive valence axis, representing positive emotions, and calmness

Table 5: Accuracy results of the negative and nonnegative task classification experiments.

Model	Train (accuracy) (%)	Test (accuracy) (%)
Logistic regression	78.31	77.78
SVM	68.67	66.67
Random forest	77.11	75.00
XGBoost	77.11	75.00
TabNet	67.47	66.67
SimAm-TabNet	61.45	61.11
Logistic regression + GAN	76.97	74.91
SVM + GAN	66.83	64.37
Random forest + GAN	76.49	75.99
XGBoost + GAN	78.32	77.45
TabNet + GAN	80.67	80.61
SimAm-TabNet + GAN	81.97	81.96

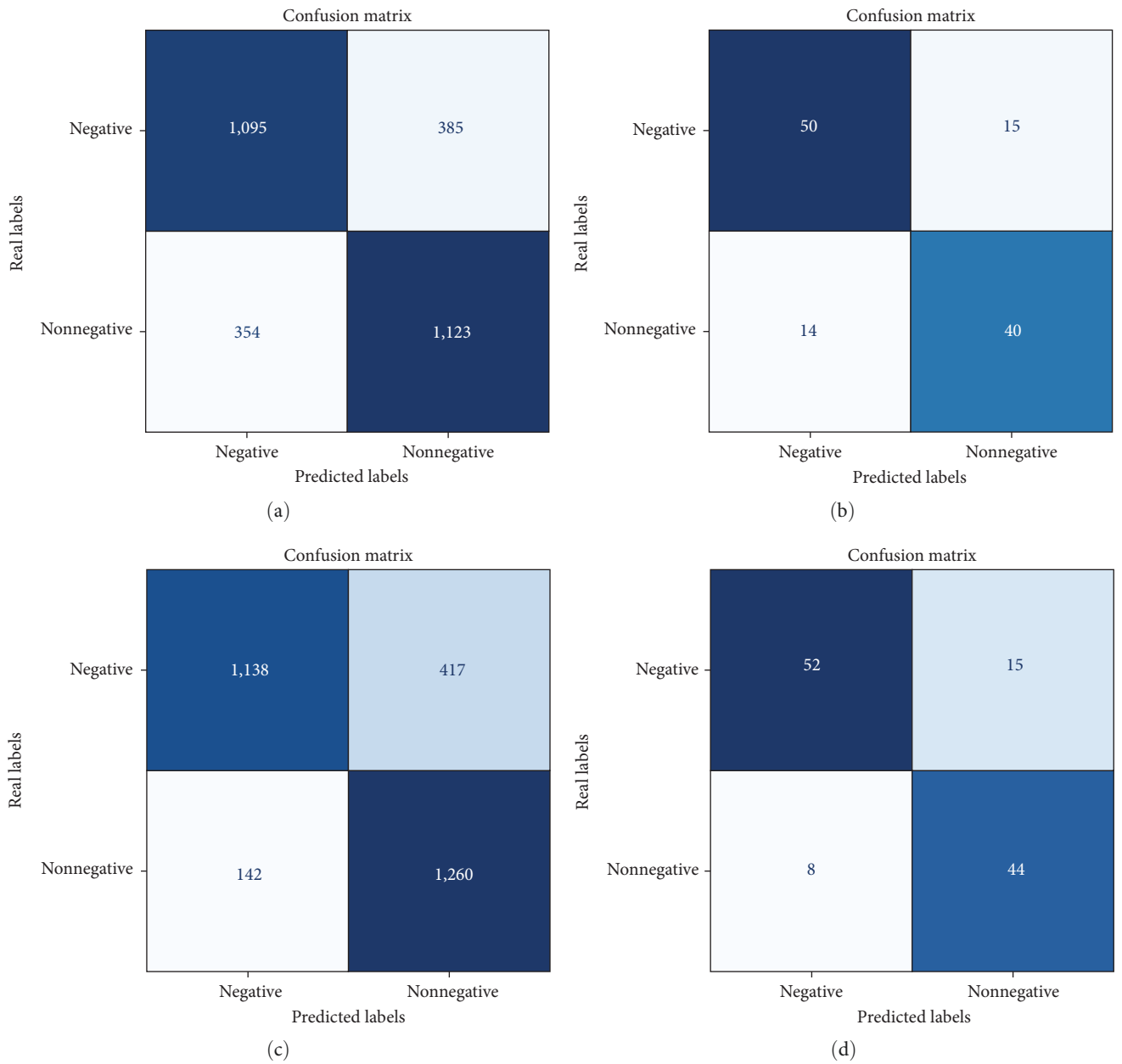

Figure 10: Detailed confusion matrix information: (a) the synthetic samples predicted by the logistic regression model trained with data generated by generative adversarial networks (GANs); (b) the original samples predicted by the logistic regression model trained with data generated by GANs; (c) the synthetic samples predicted by the SimAm-TabNet model trained with data generated by GANs; (d) the original samples predicted by the SimAm-TabNet model trained with data generated by GANs.

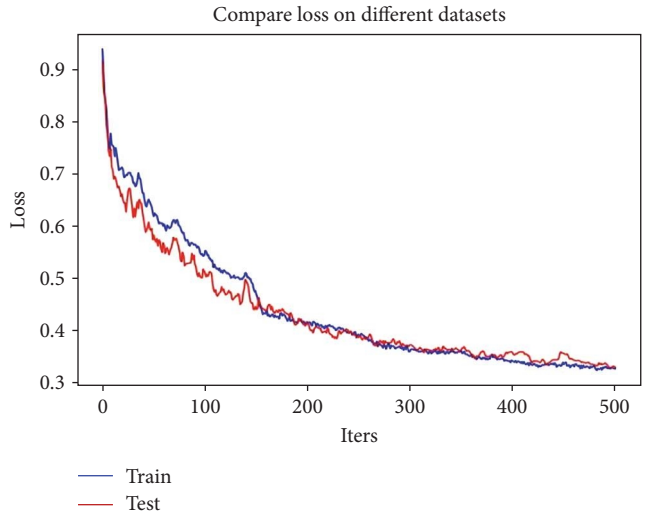

Figure 11: The training and testing loss curves of SimAm-TabNet.

Table 6: Accuracy and Kappa results of models predicting the synthetic samples and the original samples in negative and nonnegative classification experiment.

Model	The synthetic samples (accuracy/Kappa)		The original samples (accuracy/Kappa)	
Logistic regression	75.00%	0.49	75.63%	0.43
SVM	66.82%	0.38	66.39%	0.32
Random forest	76.50%	0.51	73.95%	0.44
XGBoost	76.60%	0.51	76.47%	0.49
Tabnet	78.32%	0.58	78.99%	0.59
SimAm-TabNet	81.10%	0.63	80.67%	0.62

also represents nonnegative emotions. The distinction between these two emotions is more pronounced and is reflected in handwriting.

5. Conclusions

In this study, a combination of audio-visual stimuli was used to induce happy and sad emotions in 44 participants. The collected handwriting features were organized into task samples. Machine learning models, the original TabNet deep learning model, and the TabNet deep learning model with the SimAm attention mechanism were employed for recognizing negative and nonnegative emotions. By using the CTAB-GAN, the quantity of task samples was increased, enhancing the generalization ability and reliability of the emotion recognition models. The recognition accuracy of task samples improved by 4.18%. Additionally, the incorporation of the SimAm attention mechanism led to a 1.35% improvement in classification accuracy. Ultimately, the recognition accuracy reached 80.67% on the original dataset. The findings indicate that negative and nonnegative emotions exhibit distinguishable characteristics in handwriting, and utilizing the handwriting features of writers effectively enables the recognition of negative and nonnegative emotional states.

The classification results of the emotion recognition scheme are not sufficiently high. Individual differences significantly influence emotion recognition, with higher accuracy observed for recognizing different emotions within the same individual compared to different individuals. It is recommended to increase data collection to optimize the classification models and consider integrating physiological signals with emotion-related information such as speech and facial expressions to form a multifeature fusion emotion recognition scheme, thereby improving accuracy and robustness.

Data Availability

Data will be made available upon reasonable request.

Ethical Approval

All the data are in line with the moral and ethical norms.

Consent

All the data are informed and unified by the participants.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' Contributions

Hengnian Qi was responsible for conceptualization, supervision, and project administration. Keke Jia and Gang Zeng were responsible for writing—reviewing and editing, software, investigation, data curation, and writing—original draft preparation. Chu Zhang was responsible for writing—reviewing and editing, and visualization. Mengxia Li was responsible for methodology and validation. Qing Lang was responsible for conceptualization, formal analysis, resources, and funding acquisition. Xiaoping Wu was responsible for visualization and validation. Lingxuan Wang was responsible for investigation and resources. Hengnian Qi, Gang Zeng, and Keke Jia contributed equally to this work.

Acknowledgments

This work was supported by the Chaomi S&T Company Cooperation Project (grant number: HK16003), the Postgraduate Research and Innovation Project of Huzhou University (grant number: 2022KYCX39), and the Zhejiang Key R&D Plan (grant number: 2017C03047).

References

[1] K. S. Rook, "Emotional health and positive versus negative social exchanges: a daily diary analysis," *Applied Developmental Science*, vol. 5, no. 2, pp. 86–97, 2001.

[2] L. Shu, J. Xie, M. Yang et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, Article ID 2074, 2018.

[3] B.-J. Park, C. Yoon, E.-H. Jang, and D.-H. Kim, "Physiological signals and recognition of negative emotions," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1074–1076, IEEE, Jeju, Korea (South), 2017, October.

[4] K. Dheeraj and T. Ramakrishnudu, "Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model," *Expert Systems with Applications*, vol. 182, Article ID 115265, 2021.

[5] Y. L. Murphey, M. A. Masrur, Z. H. Chen, and B. Zhang, "Model-based fault diagnosis in electric drives using machine learning," *IEEE/ASME Transactions on Mechatronics*, vol. 11, no. 3, pp. 290–303, 2006.

[6] Z. Cheng, L. Shu, J. Xie, and C. L. P. Chen, "A novel ECG-based real-time detection method of negative emotions in wearable applications," in *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pp. 296–301, IEEE, Shenzhen, China, 2017, December.

[7] M. Honda, H. Tanaka, S. Sakti, and S. Nakamura, "Detecting suppression of negative emotion by time series change of cerebral blood flow using fNIRS," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 398–401, IEEE, Las Vegas, NV, USA, 2018, March.

[8] H. E. S. Said, T. N. Tan, and K. D. Baker, "Personal identification based on handwriting," in *Pattern Recognition*, vol. 33, no. 1, pp. 149–160, 2000.

[9] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco, "EMOTHAW: a novel database for emotional state recognition from handwriting and drawing," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 273–284, 2017.

[10] Z. Sen and Z. Jing-Xiao, "A study on the influence of writer's emotional states on the characteristics of handwriting," *Journal of Criminal Investigation Police University of China*, 2019.

[11] P. Greasley, "Handwriting analysis and personality assessment: the creative use of analogy, symbolism, and metaphor," *European Psychologist*, vol. 5, no. 1, pp. 44–51, 2000.

[12] P. Singh, "General characteristics of handwriting and its psychological importance," *Cognizance Journal of Multidisciplinary Studies*, vol. 2, no. 2, pp. 1–4, 2022.

[13] H. N. Champa and K. R. AnandaKumar, "Automated human behavior prediction through handwriting analysis," in *2010 First International Conference on Integrated Intelligent Computing*, pp. 160–165, IEEE, Bangalore, India, 2010, August.

[14] B. Uğurlu, R. Kandemir, A. Carus, and E. Abay, "An expert system for determining the emotional change on a critical event using handwriting features," *TEM Journal*, vol. 5, no. 4, pp. 480–486, 2016.

[15] Y. B. Ayzeren, M. Erbilek, and E. Çelebi, "Emotional state prediction from online handwriting and signature biometrics," *IEEE Access*, vol. 7, pp. 164759–164774, 2019.

[16] K. Chaudhari and A. Thakkar, "Survey on handwriting-based personality trait identification," *Expert Systems with Applications*, vol. 124, pp. 282–308, 2019.

[17] J. Fisher, A. Maredia, A. Nixon, N. Williams, and J. Leet, "Identifying personality traits, and especially traits resulting in violent behavior through automatic handwriting analysis," in *Proceedings of Student-Faculty Research Day*, CSIS, Pace University, D6, 2012, May.

[18] M. Gavrilescu and N. Vizireanu, "Predicting the Big Five personality traits from handwriting," *EURASIP Journal on Image and Video Processing*, vol. 2018, Article ID 57, 2018.

[19] C. Taleb, M. Khachab, C. Mokbel, and L. Likforman-Sulem, "Feature selection for an improved Parkinson's disease identification based on handwriting," in *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, pp. 52–56, IEEE, Nancy, France, 2017, April.

[20] A. Esposito, T. Amorese, M. Buonanno et al., "Handwriting and drawing features for detecting personality traits," in *2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 79–84, IEEE, Naples, Italy, 2019, October.

[21] S. Al Maadeed and A. Hassaine, "Automatic prediction of age, gender, and nationality in offline handwriting," *EURASIP Journal on Image and Video Processing*, vol. 2014, Article ID 10, 2014.

[22] M. A. Mostafa, M. Al-Qurishi, and H. I. Mathkour, "Towards personality classification through Arabic handwriting analysis," in *Research & Innovation Forum 2019: Technology, Innovation, Education, and their Social Impact 1*, pp. 557–565, Springer International Publishing, 2019.

[23] J. A. Nolazco-Flores, M. Faundez-Zanuy, O. A. Velázquez-Flores, G. Cordasco, and A. Esposito, "Emotional state recognition performance improvement on a handwriting and drawing task," *IEEE Access*, vol. 9, pp. 28496–28504, 2021.

[24] S. Kedar, V. Nair, and S. Kulkarni, "Personality identification through handwriting analysis: a review," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 5, no. 1, pp. 548–556, 2015.

[25] G. Cordasco, F. Scibelli, M. Faundez-Zanuy, L. Likforman-Sulem, and A. Esposito, "Handwriting and drawing features for detecting negative moods," in *Quantifying and Processing Biomedical and Behavioral Signals*, vol. 27, pp. 73–86, Springer, Cham, 2019.

[26] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films:

a new tool for emotion researchers," *Cognition and Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
[27] P. Zheng, C.-H. Liu, and G.-L. Yu, "An overview of mood-induction methods," *Advances in Psychological Science*, vol. 20, no. 1, pp. 45–55, 2012.
[28] L. Fang, Z. Zhaohong, and B. Xuejun, "The duration of happiness and sadness induced by emotional film editing," *Studies of Psychology and Behavior*, vol. 7, no. 1, Article ID 32, 2009.
[29] X.-J. Chen, C.-G. Wang, Y.-H. Li, and N. Sui, "Psychophysiological and self-reported responses in individuals with methamphetamine use disorder exposed to emotional video stimuli," *International Journal of Psychophysiology*, vol. 133, pp. 50–54, 2018.
[30] X. Yunzi and Y. Ze, "A comparative study on the validity of different mood induction procedures (MIPs)," *Studies of Psychology and Behavior*, vol. 14, no. 5, pp. 591–599, 2016.
[31] D. Zhang, B. Wan, and D. Ming, "Research progress on emotion recognition based on physiological signals," *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*, vol. 32, no. 1, pp. 229–234, 2015.
[32] M. Fairhurst, M. Erbilek, and C. Li, "Enhancing the forensic value of handwriting using emotion prediction," in *2nd International Workshop on Biometrics and Forensics*, pp. 1–6, IEEE, Valletta, Malta, 2014, March.
[33] N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim, "Data synthesis based on generative adversarial networks," *Proceedings of the VLDB Endowment*, vol. 11, no. 10, pp. 1071–1083, 2018.
[34] Z. Zhao, A. Kunar, R. Birke, and L. Y. Chen, "CTAB-GAN: effective table data synthesizing," in *Asian Conference on Machine Learning*, pp. 97–112, PMLR, 2021.
[35] M. Esmaeilpour, N. Chaalia, A. Abusitta, F. X. Devailly, W. Maazoun, and P. Cardinal, "Bi-discriminator class-conditional tabular GAN," 2021.
[36] S.Ö. Arik and T. Pfister, "TabNet: attentive interpretable tabular learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 8, pp. 6679–6687, 2021.
[37] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "SimAM: a simple, parameter-free attention module for convolutional neural networks," in *Proceedings of the 38th International Conference on Machine Learning*, pp. 11863–11874, PMLR, 2021, July.
[38] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[40] K. Machová, M. Szabóova, J. Paralič, and J. Mičko, "Detection of emotion by text analysis using machine learning," *Frontiers in Psychology*, vol. 14, Article ID 1190326, 2023.
[41] A. U. Rahman and Z. Halim, "Identifying dominant emotional state using handwriting and drawing samples by fusing features," *Applied Intelligence*, vol. 53, no. 3, pp. 2798–2814, 2023.