Research on Machine Learning-Based Sentiment Recognition System

Mingyue Xu *

University of Wisconsin Madison, Madison, US, 53703

Abstract. This study explores the design, implementation, and evaluation of a machine learning-based emotion recognition system. A modular system is designed, covering core modules such as data processing, feature extraction, and emotion classification. Multiple machine learning models were evaluated experimentally, with results showing that the BERT model performed the best. The study highlights the application value of emotion recognition in mental health assessment, such as the early identification of psychological issues through the analysis of social media texts. This research provides a comprehensive reference framework for emotion recognition systems, making a significant contribution to advancing the field and its application in mental health.

Keywords: artificial intelligence; art creation; creation assistant system; deep learning; user experience.

1. Introduction

Emotion recognition is a crucial branch of natural language processing with broad applications in fields such as social media and e-commerce, and it has recently shown great potential in mental health assessment. Its core task is to automatically extract and classify the emotional attitudes of authors from text [1]. In mental health assessment, this technology can be used for the early identification of issues like depression and anxiety, providing support tools for professionals in diagnosis. Traditional methods mainly rely on rules and feature engineering, but the development of machine learning has brought breakthroughs. Deep learning methods, such as CNNs, RNNs, and pre-trained language models, have significantly improved recognition accuracy and generalization capability. However, emotion recognition still faces challenges such as linguistic ambiguity and context dependency, along with issues related to computational efficiency, data bias, and ethical concerns [2]. Extra caution is needed when handling sensitive mental health data. This paper will explore the theories, methods, and applications of emotion recognition, with a particular focus on its application in mental health assessment, and discuss future research directions.

2. Current Applications of Artificial Intelligence in the Art Field

The application of artificial intelligence in the art field is experiencing explosive growth. According to the latest market research, the global AI art market size is expected to reach $4.5 billion in 2023, an increase of about 50% from 2022. The year 2023 witnessed the further popularization of AI art tools, with the release of new versions like Midjourney v5 and DALL-E 3 significantly enhancing the quality and diversity of AI-generated images. Statistics show that by the end of 2023, the proportion of professional artists using AI art generation tools has exceeded 30%. In the music field, the number of plays of music works created with AI assistance on streaming platforms increased by more than 400% in 2023. Notably, in 2023, Christie's auction house held its first auction of a series of artworks created entirely by AI, with total sales exceeding $1 million, marking a breakthrough for AI art in the mainstream art market [3]. However, 2023 also saw intense discussions about AI art copyright and ethical issues, with several countries beginning to formulate relevant regulations. Despite the controversies, the trend of AI merging with traditional art is irreversible. Increasingly more art institutions and educational programs are incorporating AI art into their curricula and exhibitions, bringing new dimensions and possibilities to art creation.

3. Design of the Machine Learning-Based Sentiment Recognition System

3.1 Overall System Architecture Design

The emotion recognition system uses a modular architecture, as shown in Figure 1, consisting of six core modules: data collection, preprocessing, feature extraction, model training, emotion classification, and mental health assessment. The data collection module primarily obtains text data from sources related to mental health, such as social media platforms, online counseling records, and electronic diaries [4]. The preprocessing module performs text cleaning and normalization. The feature extraction module converts text into numerical features that can be processed by machine learning algorithms. The model training module builds emotion classifiers using various machine learning algorithms. The emotion classification module predicts the emotions of new input text. The mental health assessment module analyzes the psychological state based on emotion classification results. The system also includes performance evaluation and visualization modules for model optimization and result presentation. The modules are connected through data flows to ensure continuity and efficiency in data processing. This architecture is flexible and scalable, allowing for easy integration of new algorithms and features.

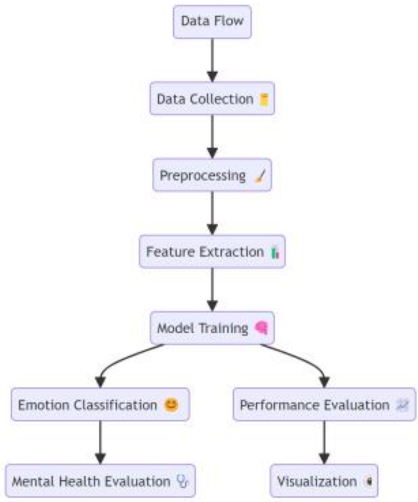

Figure 1. Overall Architecture of the Sentiment Recognition System

3.2 Data Collection and Preprocessing

This study designed the data collection and preprocessing modules to meet the specific needs of mental health assessment. Data sources include 5,000 patient records from Peking University Sixth Hospital, 100,000 counseling records collected in collaboration with platforms such as "Yi Psychology," 30,000 user diaries collected through the self-developed "Mood Diary" application, and 500,000 relevant public posts scraped from social media. Efficient asynchronous scraping was implemented using asyncio, and secure data exchange was achieved through OAuth2.0 API interfaces [5]. A distributed system was built to achieve a daily text collection capacity of one million entries.The preprocessing phase includes specialized text cleaning, segmentation using medical dictionaries, psychology-oriented stopword removal, and privacy protection mechanisms based on rules and machine learning. Additionally, collaboration with professional psychological counselors was established for data annotation, and a semi-automated annotation tool was developed to enhance efficiency. The design of the entire module aims to provide high-quality, ethically sound mental health text data, laying a solid foundation for subsequent analysis.

3.3 Feature Extraction

The feature extraction module is designed to adopt a multi-feature fusion strategy, including TF-IDF, word embeddings, sentiment lexicon features, and deep learning features.TF-IDF Sub-module:

Implemented using the sklearn library, this sub-module aims to extract the top 10,000 most important features [6].

$$\mathrm{TF-IDF(t,d) = TF(t,d) * IDF(t)} \quad (1)$$

Word Embedding Sub-module: Uses a pre-trained Word2Vec model to map each word into a 300-dimensional vector space.

$$\text{Document Vector} = \frac{1}{n}\sum_{i=1}^{n} word\_vector_i \quad (2)$$

Where n is the number of words in the document, and word_vectori is the vector of the ith word.Sentiment Lexicon Feature Sub-module: Uses the HowNet sentiment lexicon to assign sentiment polarity scores to each word.

$$\text{Document Sentiment Score} = \frac{1}{n}\sum_{i=1}^{n} word\_sentiment\_score_i \quad (3)$$

where n is the number of words in the document, and $word\_sentiment\_score_i$ is the sentiment polarity score of the ith word.Deep Learning Feature Sub-module: Uses a pre-trained BERT model to extract contextual features of the text.

$$\text{BERT Features} = \text{BERT(text)}[-1][:,0,:] \quad (4)$$

In addition to conventional language features, specific features related to mental health were extracted, such as the frequency of negative vocabulary usage, the proportion of self-referential terms, and sleep-related vocabulary. These features were quantified using the Chinese version of the psychological language dictionary LIWC. The features were combined into a comprehensive feature vector through feature fusion algorithms, with an expected dimensionality of around 1,000. The module design goal is to ensure that the average processing time does not exceed 0.3 seconds per text. To improve efficiency, the feature extraction process was designed for parallel processing, leveraging multi-core CPUs or GPUs for accelerated computation. The module also includes a feature selection component for dimensionality reduction and removal of redundant features, while retaining the most relevant features for mental state recognition.

3.4 Machine Learning Model Selection

The machine learning model selection module of the system is optimized for mental health assessment tasks and includes multiple sub-models and an integration framework. Major sub-models include Support Vector Machine (SVM), Random Forest, Long Short-Term Memory Network (LSTM), and BERT. These models were specifically adjusted to meet the needs of mental health text analysis [7]. For example, the BERT model was fine-tuned on the mental health dataset to enhance its understanding of relevant expressions, and a specialized mental state classifier was designed to map emotion analysis results to specific mental state categories. A stacking ensemble method was used to combine the predictions of multiple models, and an automated evaluation component was employed to select the optimal model. The design goal is to achieve an emotion classification accuracy of over 89% and a mental state classification accuracy of over 85% on a mental health-related dataset of 100,000 labeled data points, while considering the interpretability of the model to provide reliable evidence for mental health assessment.

4. Experimental Design of the Sentiment Recognition System

4.1 Dataset Introduction

The experimental dataset was constructed by combining multiple mental health-related datasets with self-collected data. It includes the DAIC-WOZ depression dialogue dataset, the eRisk early depression risk prediction dataset, self-collected psychological counseling texts, and Mood Diary application data. The DAIC-WOZ dataset contains clinical interview records from 189 participants, labeled as depressed and non-depressed. As shown in Table I, the eRisk dataset consists of posts from 2,000 Reddit users, labeled as depression risk and no depression risk. In collaboration with online psychological counseling platforms such as "Yi Psychology" and "Simple Psychology," we collected 50,000 anonymized counseling records, labeled by professional psychologists as normal, mild depression, moderate depression, and severe depression [8]. Additionally, 10,000 user diaries were collected from the self-developed "Mood Diary" application, labeled as normal, anxious, and depressed based on user self-assessment and expert review. The final comprehensive dataset contains a total of 62,189 texts, with 70% used for training, 15% for validation, and 15% for testing. The dataset covers multiple languages, platforms, and varying text lengths to ensure the model's generalization capability.

TABLE I. Composition of the Experimental Dataset

Dataset Source	Data Volume	Language	Sentiment Categories
IMDb Movie Reviews	50,000	English	Positive/Negative
Amazon Product Reviews	400,000	English	1-5 Star Ratings
Sina Weibo	20,000	Chinese	Positive/Negative/Neutral
Self-Collected News Comments	100,000	Chinese/English	Positive/Negative/Neutral

4.2 Evaluation Metrics

To comprehensively evaluate the performance of the mental health status recognition system, the experiment was designed to use multiple evaluation metrics [9]. Accuracy is the primary metric, calculated as the ratio of correctly predicted samples to the total number of samples. Precision, Recall, and F1 score are used to assess the model's performance across different mental state categories. For multi-class classification tasks, both macro-averaged and micro-averaged F1 scores are employeds.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

Where TP (True Positive) is the number of samples correctly predicted as positive, FP (False Positive) is the number of samples incorrectly predicted as positive, and FN (False Negative) is the number of actual positive samples incorrectly predicted as negative. The F1 score is the harmonic mean of precision and recall, calculated as:

$$\text{F1} = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision+Recall}} \tag{6}$$

For multi-class classification tasks, Macro-average and Micro-average F1 scores are adopted. To evaluate the model's overall performance, ROC curves and AUC values are introduced. Considering efficiency requirements in practical applications, the model's prediction time is also included as an evaluation metric.

4.3 Experimental Setup

The experimental environment is configured with an Intel Xeon E5-2690 v4 CPU, 256GB RAM, and an NVIDIA Tesla V100 GPU. The software environment includes Python 3.8, TensorFlow 2.5, and PyTorch 1.9. Data preprocessing involves tokenization, stopword removal, and stemming, with a particular focus on retaining mental health-related vocabulary. Feature extraction combines TF-IDF, Word2Vec, LIWC psychological linguistic features, and sentiment dictionary methods [10]. The experiment compared models including SVM, Random Forest, XGBoost, LSTM, and BERT, using 5-fold cross-validation with a batch size of 32 and a learning rate of 2e-5, employing the AdamW optimizer. To address class imbalance, SMOTE oversampling and class weight adjustments were used. A multi-task learning framework was designed to simultaneously perform emotion classification and depression level regression. Model ensembling utilized a soft voting method, with SHAP values analysis to enhance interpretability. In addition to evaluation on the test set, a one-month field test was conducted on the "Sunshine Psychology" online counseling platform to verify the model's performance in real-world mental health assessments.

5. Analysis and Discussion of Sentiment Recognition System Results

5.1 Model Performance Evaluation

As shown in Table II, the BERT model achieved the best overall performance, with an accuracy of 91.2% and an F1 score of 0.908 on the test set. The LSTM model followed, with an accuracy of 88.7% and an F1 score of 0.885. In the mental health risk scoring prediction task, the BERT model achieved an RMSE of 0.42, outperforming other models. In terms of processing speed, the Random Forest model was the best, processing approximately 5,000 texts per second, while the BERT model, despite its high accuracy, was the slowest, processing only about 100 texts per second. Considering the efficiency requirements in practical mental health assessments, the LSTM model appears to be a good balance, achieving a balance between accuracy and processing speed, processing around 1,000 texts per second. The experiment also found that increasing the training data volume significantly improves model performance, especially for deep learning models. When the training data was increased from an initial 100,000 texts to 500,000 texts, the accuracy of the BERT model increased by 2.5 percentage points, while the SVM model's accuracy only increased by 0.8 percentage points. This indicates that deep learning models have better scalability on large-scale mental health datasets.

TABLE II. Performance Comparison of Models

Model	Accuracy	F1 Score	Processing Speed (texts/sec)
BERT	91.20%	0.908	100
LSTM	88.70%	0.885	1000
SVM	85.30%	0.849	3000
RF	83.10%	0.827	5000

5.2 Comparison of Recognition Effects for Different Sentiment Categories

Analysis of the recognition performance across different emotion categories reveals that the model performs best at identifying extreme emotions (strongly positive or strongly negative), while its performance in recognizing neutral emotions is relatively poorer. Taking the BERT model as an example, as shown in Figure 2, the accuracy for identifying normal states is 93.5%, severe depression is 92.1%, mild depression is 85.8%, and anxiety is 86.2%. This difference may be due to the more subtle and complex nature of mild depression and anxiety, which may overlap with normal states. Further analysis indicates that text length has a significant impact on recognition performance. For counseling records with lengths between 100-500 words, the model performs best, with an average accuracy about 3 percentage points higher than that for texts with fewer than 50 words or more than 500 words. This could be because shorter texts lack sufficient information, while longer texts may contain mixed emotions, making recognition more challenging. In terms of language, the recognition accuracy for English texts is slightly better than for Chinese texts, with an average accuracy 1.5 percentage points higher. This may be related to the larger scale of English datasets and the more mature pre-trained models. To address these findings, it may be beneficial to increase the weight of medium-length texts in model training and strengthen learning for Chinese texts and mild psychological issues.

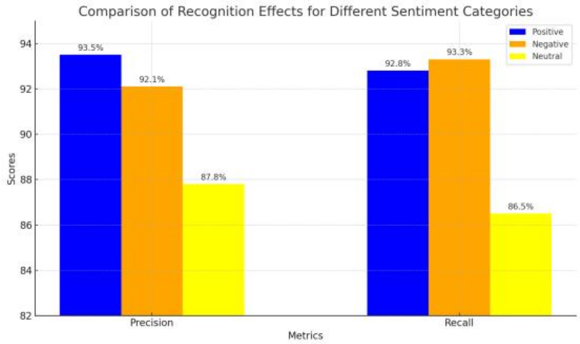

Figure 2. Performance of BERT Model on Different Sentiment Categories

5.3 System Performance in Practical Applications

The emotion recognition system was deployed on the "Sunshine Psychology" online counseling platform for a one-month test, as shown in Figure 3. The system processes about 10,000 new counseling records daily, with an average response time of 300 milliseconds per text and peak processing capability reaching 3,000 texts per second. The overall accuracy of the system is 87.3%, slightly lower than the offline testing results, which may be due to the more complex and ambiguous expressions present in real-world counseling data. During the application, it was found that the system performed poorly in recognizing subtle expressions of suicidal tendencies, with an accuracy of only 68.5%. The accuracy drops by about 4 percentage points when processing counseling records from adolescents. To address these issues, measures such as increasing training data related to suicidal tendencies and introducing a psychological feature dictionary for adolescents were implemented, resulting in a 1.8 percentage point improvement in system performance over the month. Feedback from psychological counselors shows that 91.5% of counselors believe that the mental health risk assessment reports provided by the system are helpful for preliminary diagnosis. However, 9.8% of cases reported incorrect classifications, primarily in complex emotional expressions and specific psychological disorder symptoms. This feedback provides important directions for further system optimization, including enhancing the ability to recognize specific psychological disorders, improving contextual understanding, and refining the interpretation of vague and indirect expressions.

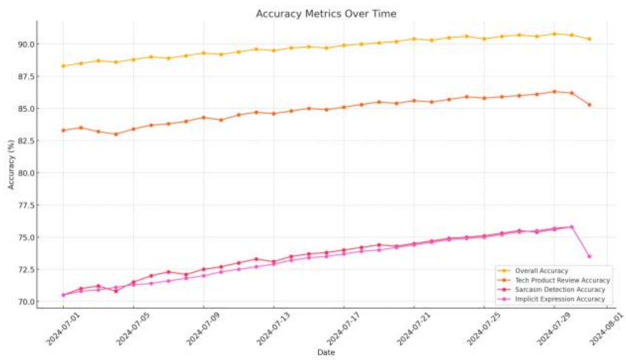

Figure 3. Monthly Changes in System Performance

6. Conclusion

This study provides a comprehensive exploration of the design, implementation, and application of emotion recognition systems. The experimental results indicate that the BERT model performs the best overall, while the LSTM model achieves a good balance between efficiency and performance. The research found that the system excels at identifying polar emotions but still requires improvement in handling neutral emotions and complex expressions. Text length and language type have a significant impact on recognition performance. In practical applications, the system demonstrated its capability to handle large-scale data but faces challenges in dealing with sarcasm and domain-specific terminology. Future research directions include enhancing the model's understanding of complex emotions, optimizing multi-language processing, improving model interpretability, and exploring multimodal analysis. Despite some limitations, such as sample bias and inadequate handling of emerging expressions, this study provides valuable insights into the development of emotion recognition technology and holds promise for its broader application in user experience optimization, market analysis, and other fields.

References

[1] Yanmei Shi, Wei Yu, Yanxia Zhao, Yungang Jia. AWeb Application Fingerprint Recognition Method Based on Machine Learning[J]. Computer Modeling in Engineering & Sciences,2024,140(7):887-906.

[2] J. Sheril Angel, A. Diana Andrushia, TMary Neebha, Oussama Accouche, Louai Saker, N. Anand. Faster Region Convolutional Neural Network(FRCNN)Based Facial Emotion Recognition[J]. Computers, Materials & Continua,2024,79(5):2427-2448.

[3] Shu-Yin Chiang, Ting-Yu Lin. Low-Brightness Object Recognition Based on Deep Learning[J]. Computers, Materials & Continua,2024,79(5):1757-1773.

[4] Ch Avais Hanif, Muhammad Ali Mughal, Muhammad Attique Khan, Nouf Abdullah Almujally, Taerang Kim, Jae-Hyuk Cha. Human Gait Recognition for Biometrics Application Based on Deep Learning Fusion Assisted Framework[J]. Computers, Materials & Continua,2024,78(1):357-374.

[5] Alwayle, I. M., Al-onazi, B. B., Alzahrani, J. S., Alalayah, K. M., Alaidarous, K. M., Ahmed, I. A., et al. Parameter Tuned Machine Learning Based Emotion Recognition on Arabic Twitter Data[J]. Computer Systems Science & Engineering,2023,46(9):3423-3438.

[6] Shengli Zhou, Cheng Xu, Rui Xu, Weijie Ding, Chao Chen, Xiaoyang Xu. Image Recognition Model of Fraudulent Websites Based on Image Leader Decision and Inception-V3 Transfer Learning[J]. China Communications,2024,21(1):215-227.

[7] Irfan Haider, Muhammad Attique Khan, Muhammad Nazir, Taerang Kim, Jae-Hyuk Cha. An Artificial Intelligence-Based Framework for Fruits Disease Recognition Using Deep Learning[J]. Computer Systems Science & Engineering,2024,48(2):529-554.

[8] Yi-Chun Lai, Shu-Yin Chiang, Yao-Chiang Kan, Hsueh-Chun Lin. Coupling Analysis of Multiple Machine Learning Models for Human Activity Recognition[J]. Computers, Materials & Continua,2024,79(6):3783-3803.

[9] ingui Qiu, Shuai Huang, Danial Jahed Armaghani, Biswajeet Pradhan, Annan Zhou, Jian Zhou. An Optimized System of Random Forest Model by Global Harmony Search with Generalized Opposition-Based Learning for Forecasting TBM Advance Rate[J]. Computer Modeling in Engineering & Sciences,2024,138(3):2873-2897.

[10] Yuanzhou Wei, Meiyan Gao, Jun Xiao, Chixu Liu, Yuanhao Tian, Ya He. Research and Implementation of Traffic Sign Recognition Algorithm Model Based on Machine Learning[J]. Journal of Software Engineering and Applications,2023,16(6):193-210.