Proceedings of the 5th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/100/2025.21191

MoodType - An Emotion Monitoring App Based on the Mobert Model

Zetong Shao[1,a,*], Peijun Tu[2,b], Yutong Gao[3,c], Yaqi Wang[4,d]

[1]*School of Information, Shanghai Ocean University, Shanghai, 201306, China*
[2]*School of Information Engineering, Wuhan University of Technology, Wuhan, 430070, China*
[3]*Qingdao No.58 Highschool of Shandong Province, Qingdao, 266000, China*
[4]*BASIS International School Park Lane Harbor, Shenzhen, 516082, China*
a. Shaozetong@163.com, b. tupeijun098@gmail.com, c. ivygao071230@outlook.com, d. Lucywang128@163.com
**corresponding author*

Abstract: Nowadays, psychological diseases can hardly be detected due to the lack of willingness to get help from professional doctors. With the appearance of smartphones, mobile applications offered a promising future for the auto-monitoring of psychological diseases. In this article, we introduced MoodType, an interactive diary application that offers related mood suggestions based on the input of users. MoodType can classify six different mood types based on the diary the user wrote in this application. Moreover, it offers white noise and pure music for users to release pressure when they are writing the diary. Our data collection comes from more than 20,000 comments on Twitter (Now X), achieving a 0.937 F1-score for the six different moods. In the process of a week's user experience stage, most of the participants held a high evaluation of both its mood analyzing and interactive experience, showing the reliability of the MoodType.

Keywords: mental health, mobile health, emotion monitoring, deep learning.

1. Introduction

Psychological diseases are recognized as the main issues in the modern society. Research shows that Major Depressive Disorder (MMD) and Generalized anxiety disorder (GAD) are the main contributors to disability and early death [1]. It is estimated that more than 300 million people around the world are suffering from MMD, which accounts for 4.4% of the world population.

Therefore, it is necessary to implement early prevention and monitoring. However, in real life, clinicians often have trouble obtaining frequent real-time monitoring. The reason that more people, especially youngsters, are in favor of monitoring their psychological state by themselves, we developed MoodType, an interactive diary application that classifies six different moods and offers related professional suggestions based on the input of users.

Our main research contributions included:

• We trained a fine-tuned pre-trained model, Mobert, for six-class emotion classification.

• MoodType identifies emotions through journaling, providing early warnings of potential psychological health issues and offering self-assessment questionnaires, enabling self-monitoring of mental health.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/100/2025.21191

• MoodType delivers professional advice based on user's specific emotional states, helping them improve their mood.

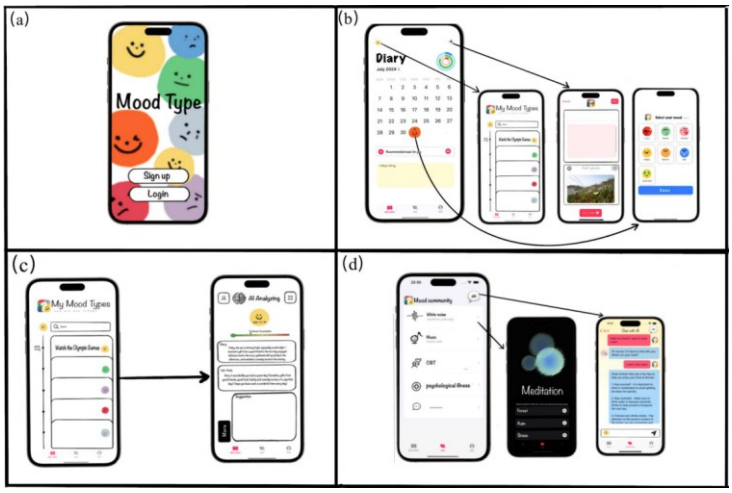
2. Related Work

The mobile phone has become an ideal psychological health monitoring platform. For example, Kwon et al took advantage of a mobile phone camera to develop a museum visitor satisfaction evaluation system that analyzes the facial expressions of visitors to detect their mood [2]; Suparatpinyo et al used the microphone of smartphones to collect voice data and developed a mobile phone app to realize the self-assessment and detection of depression [3].

Many researchers have achieved significant results in emotion classification by their fine-tuning BERT models [4]. For example, the BERT model reached an accuracy of 84.2% on the SST-5 dataset (with five sentiment classes). Similarly, Zeng Cheng et al. achieved emotion classification accuracy of 94.3%, 93.5%, and 94.8% on data from video websites (Bilibili, iQiyi, and Tencent Video) using the pre-trained ALBERT_CRNN model [5].

However, these systems only offer basic analyzes of moods, they didn't manage to detect moods automatically. The Moodtype app made a step forward and achieved a high accuracy on the Twitter Sentiment Analysis dataset (with six categories).

3. Design Architecture



(a) Splash Screen Guide Page (b) Homepage Sections (c) Diary and Related Suggestions Page (d) Mood Community Interface

Figure 1: App Design Architecture

MoodType can judge users' psychological state by detecting their diaries and offering suggestions. In this chapter, we discuss the UI design of MoodType.

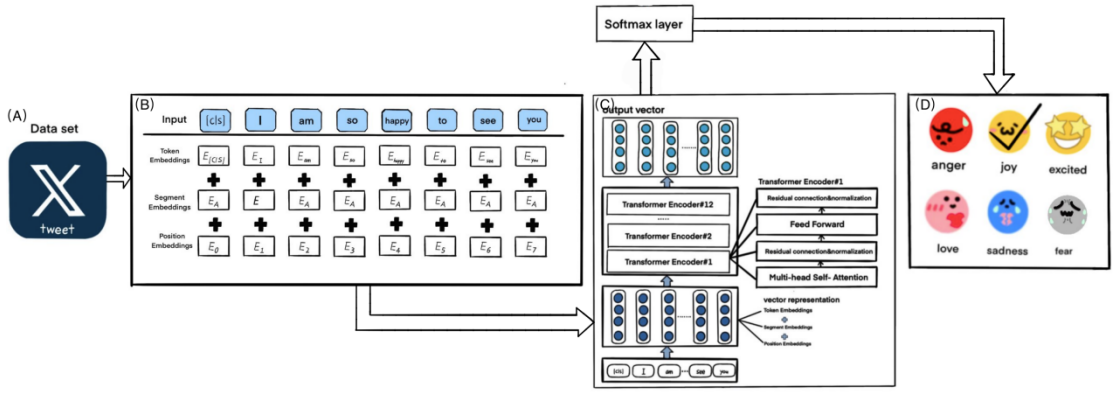
For the reason that MoodType is an app aimed at mood analysis; its design fully considers users' subjective feelings, aiming to relieve users' cognitive stress [6]. The login page (Figure 1a) used a visual design with no distractions, featuring an emotion-themed emoji to help users accept their moods and smoothly log into the app. The front page (Figure 1b) showed the app's primary functions through a modular layout, allowing users to quickly access the features they wanted. The diary interface (Figure 1c) allowed users to add diary entries by clicking the "+" icon, where they could also add images. This interface also offered options to view past moods, add current moods through the calendar, and receive music recommendations based on the user's mood state. The mood

DOI: 10.54254/2755-2721/100/2025.21191

community interface (Figure 1d) had features such as white noise, soothing music, CBT therapy [7], and basic psychological knowledge, allowing users to select services based on mood analysis.

4. Implementation

In this section, we are going to describe the details in implementing this app. Firstly, we record users' moods by writing diaries or selecting their moods. Then, we used our fine-tuned Mobert, which is based on Bert-base-uncased [8], to get good results as a baseline model. After that, we used the API of ChatGpt 4o to give the users specialized and professional advice [9]. We also made progress in 3 different aspects:(i) To improve model performance, we conducted multiple experiments on the Twitter (Now X) emotion datasets to optimize the training process; (ii) To monitor mental health issues, the software automatically triggers a warning and provides a questionnaire for self-assessment when negative emotions reach a certain threshold; (iii) MoodType also helps alleviate stress through additional features such as white noise and integration with NetEase Music's API for music recommendation.



(a)Twitter Dataset (b) Input of BERT Model (c) BERT Model Structure (d) Mood Classification Results

Figure 2: Model Architecture Flowchart

4.1. Dataset and model structure

We selected the datasets of emotion on Twitter (Figure 2a), which involve 6 basic emotions: anger, jot, excitement, love, sadness, and fear. Our dataset consists of 20,000 Twitter user messages, with 16,000 for training, 2,000 for validation, and 2,000 for testing. The classification labels include (0), (1), (2), (3), (4), and (5), corresponding to anger, joy, excitement, love, sadness, and fear, respectively. The label distribution is 29.2%, 33.5%, 8.2%, 13.5%, 12.1%, and 3.6%.

For model selection, we used the pre-trained BERT model, which showed the best performance on these datasets. The input to the BERT model consists of word embedding, segment embedding, and positional embedding [10] (see Fig. 2b). BERT assigns distinct vectors to each word and adds a [CLS] token at the beginning of the sequence [11]. Although the [CLS] token does not carry explicit semantic meaning, its output vector effectively captures the overall sentiment of the sentence, making it useful for classification tasks.

A key innovation of BERT is the use of Transformer encoders with masked multi-head attention mechanisms (see Fig. 2c). The model first integrates word embedding, segment embedding, and positional embedding and then processes these inputs sequentially through 12 Transformer encoders [12]. After undergoing linear transformation, layer normalization, dropout, and ReLU activation,

DOI: 10.54254/2755-2721/100/2025.21191

these vectors are further refined and used as inputs for subsequent encoders, ultimately producing rich responses.

4.2. Pre-trained model fine-tuning

We fine-tuned the BERT-base-uncased pre-trained model to test how different Batch Sizes and Learning Rates affect its performance. We used a linear scheduler to gradually reduce the learning rate and combined it with the Adam optimizer (β1=0.9, β2=0.999, eps=1e−8) to improve training efficiency and stability [13]. We set the training to 5 epochs and weight decay to 0.01, using the F1 score as the evaluation because it works well with datasets that are not balanced. By adjusting the Batch Size and Learning Rate, we found the best settings for our model.

4.3. Interaction with MoodType

When MoodType detects certain emotions, it tracks negative emotions. If a user has negative emotions more than five times in a week, the app sends an alert and provides a mental health self-assessment questionnaire (such as a CBT questionnaire).

The app includes an emotion list (see Fig. 2d) where users can select moods instead of typing them out, making it easier to record their feelings.

5. Result & Analysis

In this section, we will introduce our experiment result. We assessed our application in the following ways:(i) A technical evaluation of the emotion classification model performance. (ii) A survey evaluation from 13 users (aged 16-35, with a balanced gender distribution) after using the software for one week.

5.1. Technical Evaluation

We used different Batch Sizes and Learning Rates to train the model while making other initial parameters unchanged. The results are shown in Table 1. The experiment demonstrated that the highest accuracy of 93.7% was achieved when the Batch Size is 64 and a Learning Rate of 2e-5. We call this model Mobert.

Table 1: Mobert fine-tuning results on different parameter

Batch Size	Learning Rate	Accuracy	Training Loss	Test Sample per Second	F1-score
32	1e-5	0.929	0.125	518.798	0.929
64	1e-5	0.933	0.113	256.832	0.934
32	2e-5	0.930	0.121	521.621	0.932
64	2e-5	0.937	0.105	264.473	0.937
32	3e-5	0.934	0.110	529.731	0.934
64	3e-5	0.935	0.107	268.568	0.936

DOI: 10.54254/2755-2721/100/2025.21191

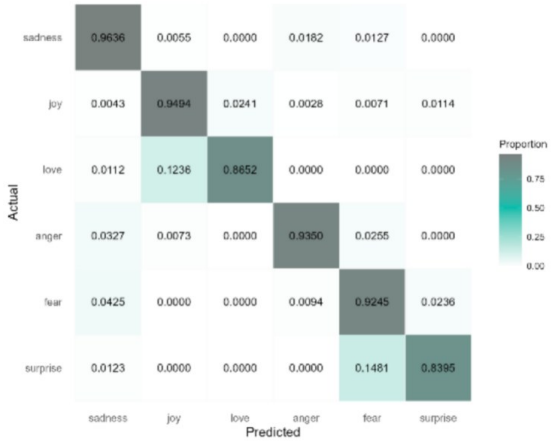


Figure 3: Confusion matrix of Twitter_emotion datasets on the test set

To better understand Mobert's performance in the datasets, we analyzed the Confusion matrix results of the test sets. Figure 3 illustrates the confusion matrix for the predicted labels versus the true labels on the Twitter dataset [15]. As observed, we can see that the confusion rate between the labels love and joy and between surprise and fear is relatively high, at 14.77% and 17.17% respectively. We found that this confusion arises mainly from two factors: (i) ambiguity in label definitions by annotators, (ii) limitations in the dataset itself, as these sentences were often short and had fewer training examples.

5.2. User Research

During the evaluation phase of the MoodType application, 13 users participated in an assessment of the app and its emotion analysis model. They focused on key aspects such as classification accuracy, sensitivity, relevance of AI-generated comments, model complexity, feedback speed, interface usability, clarity of result presentation, and overall user experience.

A total of 92.3% of users found the classification to be "very accurate" or "fairly accurate," indicating strong model performance with room for improvement. Additionally, 84.6% of users noted that the model responded sensitively to emotional changes, though further optimization is possible. Similarly, 84.6% felt that the AI-generated comments were relevant to their mood records, but they suggested that personalize areas could be enhanced. Regarding the software's performance, 92.3% of users were satisfied with the feedback speed, finding it both prompt and smooth. Furthermore, 92.3% of users appreciated the warmth and thoughtfulness of the suggestions provided by the software, reflecting general approval of its advice. As for the clarity of the result presentation, 76.9% of users found it clear and easy to understand, though some indicated that the presentation was average. And finally, 92.3% of users were satisfied with the overall fluidity of the application, which performed pretty well.

In conclusion, the users generally expressed a positive evaluation of MoodType's emotion analysis and user experience. However, there is still room for improvement in areas such as model complexity, personalization of AI comments, and clarity of result presentation. These aspects will be the focus of future development to further enhance the user experience.

6. Conclusion

We introduced MoodType, a mobile application designed for the automatic monitoring of psychological health issues. By using the fine-tuned pre-trained model Mobert, MoodType accurately classifies six common emotions expressed by users in their journals. The app not only tracks users'

DOI: 10.54254/2755-2721/100/2025.21191

emotional data and alerts users when negative emotions reach a certain degree but also offers self-assessment questionnaires for psychological health. Additionally, MoodType generates personalized professional advice based on users' specific emotional states, helping them better manage and improve their moods. The integration of deep learning-based result visualization and localization further enhances users' understanding and experience.

Acknowledgement

Zetong Shao, Peijun Tu, Yutong Gao, and Yaqi Wang contributed equally to this work and should be considered co-first authors.

References

[1] *Chodavadia P, Teo I, Poremski D, Fung DS, Finkelstein EA. Prevalence and economic burden of depression and anxiety symptoms among Singaporean adults: results from a 2022 web panel. BMC psychiatry. 2023 Feb 14;23(1):104.*

[2] *Do Hyung Kwon and Jeong Min Yu. 2024. Real-time Multi-CNN-based Emotion Recognition System for Evaluating Museum Visitors' Satisfaction. J. Comput. Cult. Herit. 17, 1, Article 15 (March 2024), 18 pages.*

[3] *Suparatpinyo, S., Soonthornphisaj, N. Smart voice recognition based on deep learning for depression diagnosis. Artif Life Robotics 28, 332–342 (2023).*

[4] *Acheampong, F.A., Nunoo-Mensah, H. & Chen, W. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artif Intell Rev 54, 5789–5829 (2021).*

[5] *Socher R, Perelygin A, WU J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Findings of the Association for Computational Linguistics, 2013: 1631-1642.*

[6] *Zeng Cheng, Wen Chaodong, Sun Yumin, Pan Li & He Peng. (2021). Emotion analysis of bullet screen text based on ALBERT-CRNN. Journal of Zhengzhou University (Science Edition) (03), 1-8.*

[7] *Sudak, D. M. (2012). Cognitive behavioral therapy for depression. Psychiatric Clinics, 35(1), 99-110.*

[8] *Turc, I., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Well-read students learn better: On the importance of pre-training compact models (arXiv preprint arXiv:1908.08962v2).*

[9] *Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (Generative pre-trained transformer)—A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. IEEE Access, 12, 54608-54649.*

[10] *Prabhu, R., & Nashappa, C. S. (2024). Multi-source word-aligned attention average pooling based bidirectional encoder representation from transformers for product review sentiment analysis. In Proceedings of the 2024 4th International Conference on Data Engineering and Communication Systems (ICDECS) (pp. 1-6). IEEE.*

[11] *He, A., & Abisado, M. (2024). Text sentiment analysis of Douban film short comments based on BERT-CNN-BiLSTM-Att model. IEEE Access, 12, 45229-45237.*

[12] *Pratama, B. T., Utami, E., & Sunyoto, A. (2019). A comparison of the use of several different resources on lexicon based Indonesian sentiment analysis on app review dataset. In 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) (pp. 282-287). IEEE.*

[13] *Tato, A., & Nkambou, R. (2018). Improving Adam optimizer.*

[14] *Gulmatico, J. S., Susa, J. A. B., Malbog, M. A. F., Acoba, A., Nipas, M. D., & Mindoro, J. N. (2022, March). SpotiPred: A machine learning approach prediction of Spotify music popularity by audio features. In 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T) (pp. 1-5). IEEE.*

[15] *Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-label confusion matrix. IEEE Access, 10, 19083-19095.*