

# TCU at ROCLING-2025 Shared Task: Leveraging LLM Embeddings and Ensemble Regression for Chinese Dimensional Sentiment Analysis

運用大型語言模型嵌入與集成迴歸技術進行中文維度情感分析

李信頡 (Hsin-Chieh Li)

Department of Medical Informatics  
Tzu Chi University  
113206105@gms.tcu.edu.tw

林紋正 (Wen-Cheng Lin)\*

Department of Medical Informatics  
Tzu Chi University  
denislin@gms.tcu.edu.tw

## 摘要

本文參加 ROCLING-2025 共享任務：針對醫學自我反思文本的中文維度情感分析。維度式情感分析 (Dimensional Sentiment Analysis, DSA) 將情緒視為連續維度，如效價 (valence, 正向至負向) 與喚醒度 (arousal, 平靜至激動)，相較傳統分類式方法提供更細膩的表徵，適用於心理健康監測與風險預測等應用。利用大型語言模型 (Large Language Models, LLMs) 作為特徵提取器生成上下文嵌入向量，並應用於支持向量迴歸 (SVR) 等迴歸模型進行效價-喚醒度預測。訓練資料採用 Chinese EmoBank 資料集 (2,954 筆通用領域樣本)，驗證資料為醫療自省文件語料資料集 (994 筆)，測試資料為醫療自省文件語料資料集 (1,541 筆)。實驗結果顯示，使用 DeepSeek 嵌入的 SVR 模型表現最佳，在此基礎上透過多模型集成學習，效能提升至 valence MAE: 0.463、arousal MAE: 0.759、valence PCC: 0.805、arousal PCC: 0.608。此方法突顯多模型融合在 DSA 於生醫情境的潛力，促進非侵入式心理健康評估工具的發展。

## Abstract

This study participates in the ROCLING-2025 shared task on Chinese dimensional sentiment analysis for medical self-reflection texts. Dimensional Sentiment Analysis (DSA) represents emotions as continuous dimensions, such as valence (positive to negative) and arousal (calm to excited), providing finer-grained representations compared to traditional categorical approaches, which are suitable for applications in mental health monitoring and risk assessment. We use large language models (LLMs) to extract contextual embedding vectors, which are then fed into regression models, such as Support Vector Regression (SVR), for valence-arousal prediction.

\*通訊作者 (Corresponding Author)

The training data consists of the Chinese EmoBank dataset (2,954 general-domain samples), the validation data is a Medical Self-Reflection Corpus Dataset (994 samples), and the test data is another Medical Self-Reflection Corpus Dataset (1,541 samples). Experimental results show that the SVR model with DeepSeek embeddings performs best. Multi-model ensemble learning further improves performance to 0.463 valence MAE, 0.759 arousal MAE, 0.805 valence PCC, and 0.608 arousal PCC. This approach shows the potential of multi-model fusion in DSA for biomedical applications, facilitating the development of non-intrusive mental health assessment tools.

關鍵字：維度式情感分析、大型語言模型、集成學習

**Keywords:** Dimensional Sentiment Analysis, Large Language Models, Ensemble Learning

## 1 引言

情感分析 (Sentiment Analysis)，又稱為意見挖掘 (Opinion Mining)，是自然語言處理 (Natural Language Processing, NLP) 領域中的核心任務之一，旨在自動辨識、提取與量化文本資料中所表達的主觀情感與態度。該技術廣泛應用於社群媒體評論分析、產品意見理解、政治輿論監測等領域，並逐漸拓展至醫療與心理健康等應用場景，用以協助病患情緒追蹤與心理風險預測 (Liu, 2012)。

情緒在心理健康評估中扮演關鍵角色，隨著數位語言資料與人工智慧技術的發展，學界與醫療實務者逐漸從日常語言中探索情緒狀態與心理健康的關聯。

研究指出，情緒解析能力 (辨識與表達細緻情緒的能力) 與心理健康密切相關 (Vishnubhotla et al., 2024)。在社群媒體上，情緒表達模糊或情緒用語貧乏的使用者，更易呈現憂

鬱、焦慮等風險，反映情緒語言的豐富程度可作為非侵入式心理健康指標。

另一方面，線上心理健康社群的語言分析發現，不同精神疾病群體展現出獨特的情緒語言特徵 (Yan et al., 2021)。例如，焦慮群體常帶有未來導向與警戒感，憂鬱群體則偏向無望與自我否定。此差異有助於辨識心理健康問題，且社群互動中的回應亦具情緒調節作用。

綜合而言，情感分析在心理健康預測、疾病風險篩檢與數位健康監控方面展現潛力，透過日常語言的情緒線索，有望發展低成本、即時性的心理健康監測工具，推動預防醫學與個人化照護。

傳統的情感分析多採分類式方法 (categorical approach)，將情緒分為「正面」、「負面」或「中性」等類別。然而，這類方法在面對語言中細膩或混合的情緒表達時，往往難以準確反映真實的心理狀態 (Mohammad, 2016)。為此，學界提出維度式情感分析 (Dimensional Sentiment Analysis, DSA) 作為替代方案，依據心理學理論將情緒視為連續空間中的向量位置，常見的表示方式包括「效價 (Valence)」、「喚醒度 (Arousal)」二個維度。(Calvo and D'Mello, 2010; Jonathan Posner and Peterson., 2005)。

ROCLING-2025 舉辦了共享任務：針對醫學自我反思文本的中文維度情感分析 (Lee et al., 2025)，在本任務中參與者需要為每位醫生的自我反思文本，在情緒效價和喚醒度維度上給出 1 到 9 的實際值評分。本文將說明我們在此任務中使用的方法與結果。

## 2 相關工作

### 2.1 早期方法：基於詞典與傳統機器學習

早期研究多依賴於情感詞典與傳統機器學習模型。情感詞典法透過預先標註好 VA 值的情感詞典 (如 ANEW, CVAW) (Warriner et al., 2013; Yang et al., 2016)，計算文本中詞語的平均 VA 分數作為整句的情感預測。傳統機器學習方法，如支持向量迴歸 (Support Vector Regression, SVR) (Drucker et al., 1997)，則將文本轉換為詞袋 (Bag-of-Words) 或 TF-IDF 等特徵向量，再進行迴歸預測 (Malandrakis et al., 2013)。

### 2.2 深度學習

以卷積神經網路 (CNN) 和循環神經網路 (RNN) 為代表的模型被廣泛應用於 DSA。CNN 擅長擷取局部特徵 (Kim, 2014)，而 LSTM 等 RNN 變體則能處理序列資訊 (Tai et al., 2015)。有研究將 CNN 與 LSTM 結合，以同時捕捉局部與全局語意 (Hasib et al.,

2023)。此外，注意力機制 (Attention Mechanism) 的引入亦顯著提升了預測效能 (Yang et al., 2016)，(林巍, 2022) 則指出多任務學習有助於捕捉維度間的關聯性。

為克服傳統「詞袋模型」(Bag-of-Words) 無法保留語序與語意的限制，詞嵌入 (Word Embedding) 技術被提出，用於將文字轉換為實數向量，以利於語義分析。早期方法如 word2vec (Mikolov et al., 2013) 與 GloVe (Pennington et al., 2014) 為每個詞生成固定向量；隨著 Transformer 架構的興起，BERT (Devlin et al., 2019) 等預訓練語言模型能提供上下文相關的「語境嵌入」(Contextual Embedding)，使同一詞在不同語境中具有不同表示，大幅提升了語意建模能力。

### 2.3 大型語言模型

大型語言模型 Large Language Models, LLMs) 是由具有大量參數的類神經網路組成的一類語言模型，如 GPT (OpenAI, 2023)、Llama (Touvron et al., 2023)、DeepSeek (Guo et al., 2025) 等，展現出驚人的語意理解與生成能力。研究者常以其作為特徵提取器，將輸出的語意向量輸入傳統機器學習模型 (如 SVR、隨機森林) 進行 VA 值預測。

### 2.4 近期方法

首先，在情感維度方面，研究逐漸引入第三維 (Dominance / 支配度) 或更多維度，以捕捉情緒控制力與社會互動等特徵。Yang 等人提出的 SCCL 模型 (Semantic Cluster-level Contrastive Learning) (Yang et al., 2023)，在對話式情緒識別中利用 VAD 三維空間結合對比學習，有效提升模型可解釋性與穩定性。

其次，Transformer 架構與參數高效微調 (Parameter-Efficient Fine-Tuning, PEFT) (Vaswani et al., 2023) 技術被廣泛應用於情緒迴歸任務。部分研究透過在預訓練模型上附加迴歸層進行微調，或採用提示式 (prompt-based) 策略，於少樣本與跨領域任務中取得良好成效 (Wawer, 2024)。

此外，對比學習 (contrastive learning) 成為強化情緒嵌入的重要手段。Hu 等人的 LaSCL (Label Semantic-Driven Contrastive Learning) (Hu et al., 2025) 利用情緒標籤語義嵌入作為語義錨點以強化情緒區分，Xie 等人的 DCLF (Dual Contrastive Learning Framework) (Xie et al., 2025) 則結合多模態與上下文對比學習，以提升特徵融合的整體效果。

在多模態融合方面，近期研究整合文本、語音、影像及生理訊號等多來源特徵，並透過跨模態注意力 (cross-attention) 與維度專屬融合

機制 (dimension-wise fusion) 改善模型一致性與穩健性。例如 DCLF 模型與 PCMDA (Parallel Contrastive Multimodal Domain Adaptation) (Li et al., 2025) 均在多模態情緒識別中展現良好成效。

最後，領域適配 (domain adaptation) 與多指標評估逐漸受到重視。Wawer (Wawer, 2024) 探討少樣本學習於情感分析中的跨域適應，顯示原型網路與相似度式學習對資料稀缺問題具優勢。同時，評估方式亦從單一誤差指標擴展至 MAE、PCC、CCC 等多重指標，以全面反映模型在連續值預測的表現。

### 3 方法

本研究以大型語言模型所提取之文本特徵作為輸入，並將其應用於多種機器學習方法進行情緒迴歸任務的建模。所選用之模型涵蓋傳統機器學習方法與深度學習方法兩大類，具體如下：

- 傳統機器學習方法：  
SVR、  
LightGBM (Ke et al., 2017)、  
XGBoost (Chen and Guestrin, 2016)、  
CatBoost (Prokhorenkova et al., 2017)。
- 深度學習方法：  
以多層神經網路為基礎之迴歸模型。

#### 3.1 資料集

訓練資料集使用 Chinese EmoBank (Lee et al., 2022; Yu et al., 2016) 內容包含 2,954 筆通用領域中文文本，涵蓋新聞、政治、飯店評論、書籍、汽車及筆記型電腦等。文本長度：平均 57.6 字元，標準差 32.6，多數集中於 51-100 字元。情感標註：Valence 平均 4.8，Arousal 平均 4.8，近似常態分佈，集中於中間值。驗證資料集：內容包含 994 筆醫療自省文本，高頻詞為醫學術語（如「病人」、「治療」）。文本長度：平均 76.5 字元，標準差 53.1，長於訓練集，呈右偏分佈。情感標註：Valence 平均 4.1，Arousal 平均 4.0，情感偏負面且喚醒度較低。測試資料集 (DSAMST)：內容包含 1541 筆醫療自省文本。所有資料皆依照維度

式情感模型 (Dimensional Sentiment Model) 進行標註，標籤包含兩個連續變數：

- 情感效價 (Valence)：表示情感的正向或負向程度，評分範圍為 1 至 9 分，數值越高代表越正面（如愉悅、滿意），數值越低則偏向負面（如悲傷、憤怒）。
- 喚醒度 (Arousal)：表示情感的強度或激動程度，評分範圍同為 1 至 9 分，數值越高代表越激動或精力充沛（如亢奮、緊張），數值越低則表示較為平靜或低落（如冷靜、無力）。

#### 3.2 評估指標

模型的性能將根據兩個指標進行評估，分別針對 Valence 和 Arousal 獨立計算：

平均絕對誤差 (Mean Absolute Error, MAE) 評估模型預測值與人工標註真實值之間的平均絕對差距。其計算公式如下：

$$MAE = \frac{1}{n} \sum_{i=1}^n |a_i - p_i| \quad (1)$$

其中， $n$  為樣本數， $a_i$  為真實值， $p_i$  為預測值。MAE 越小，表示模型誤差越低。

皮爾森相關係數 (Pearson Correlation Coefficient, PCC) 評估模型預測值與真實值之間的線性相關程度，範圍在 -1 到 1 之間。其計算公式如下：

$$PCC = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{a_i - \mu_A}{\sigma_A} \right) \left( \frac{p_i - \mu_P}{\sigma_P} \right) \quad (2)$$

其中， $\mu_A$  與  $\mu_P$  分別為真實值與預測值的平均數， $\sigma_A$  與  $\sigma_P$  為其標準差。PCC 越接近 1，表示相關性越強。

#### 3.3 情感分析模型

本研究的核心流程為：首先使用大型語言模型作為編碼器 (Encoder)，將醫學文本轉換為上下文語意嵌入向量；接著將嵌入向量輸入多種

Combination	MAE_V	MAE_A	PCC_V	PCC_A
Training	0.528	0.914	0.711	0.483
Validation	0.553	0.801	0.706	0.537
Training + half validation	0.524	0.809	0.754	0.538
Training + Validation	0.524	0.809	0.754	0.538

Table 1: 不同訓練資料組合實驗結果

Model Name	MAE_V	MAE_A	PCC_V	PCC_A	Overall Rank
XGBoost	0.526 (3)	0.857 (5)	0.724 (5)	0.457 (5)	4.50 (5)
LGBM	0.519 (1)	0.852 (4)	0.740 (3)	0.472 (4)	3.00 (3)
CatBoost	0.538 (4)	0.834 (3)	0.732 (4)	0.495 (3)	3.50 (4)
SVR	0.524 (2)	0.809 (1)	0.754 (2)	0.538 (2)	1.75 (1)
CustomResNet	0.550 (5)	0.830 (2)	0.761 (1)	0.542 (1)	2.25 (2)

Table 2: 不同機器學習模型之實驗結果比較

Encoder Name	MAE_V	MAE_A	PCC_V	PCC_A
DeepSeek-R1-0528-Qwen3-8B	0.524	0.809	0.754	0.538
DeepSeek-Prover-V1.5-RL	0.527	0.839	0.749	0.504
Llama3-TAIDE-LX-8B-Chat-Alpha1	0.529	0.827	0.743	0.518
multilingual-e5-large	0.554	0.810	0.726	0.531
multilingual-e5-large-instruct	0.523	0.807	0.742	0.539

Table 3: 不同編碼器之實驗結果比較

迴歸模型 (Regressors)，以預測效價 (Valence) 與喚醒度 (Arousal) 數值。每一次實驗 (Run) 即對應一組「Encoder + Regressor」的組合，並比較其效能表現。

**LLM Encoders** 本研究使用五種大型語言模型作為編碼器，以產生輸入文本的上下文嵌入向量。DeepSeek-R1-0528-Qwen3-8B 與 DeepSeek-Prover-V1.5-RL (Xin et al., 2024) 均屬於近年提出的 DeepSeek 系列模型，前者強調通用語理解，後者則在推理與邏輯任務上進行優化。Llama3-TAIDE-LX-8B-Chat-Alpha1 (TAIDE, 2025) 為針對中文及多語言對話任務調校的 Llama3 模型版本，具備良好的語境捕捉能力。另一方面，multilingual-e5-large 與 multilingual-e5-large-instruct (Wang et al., 2024) 則為針對檢索與語意相似度任務設計的嵌入模型，能在多語言環境下提供高效且一致的語意表示。這些編碼器能將輸入文本轉換為高維度向量表示，進一步保留與情緒相關的語意特徵，並支援後續的迴歸任務。

**Regression Models** 我們在實驗中比較了多種迴歸模型，涵蓋傳統機器學習方法與深度學習方法。支持向量迴歸 (SVR) 採用 RBF kernel，主要參數設為  $C = 10$  與  $\epsilon = 0.2$ ，能有效處理非線性分佈，其核心思想是透過高維特徵映射尋找最佳迴歸超平面。XGBoost 模型則使用  $n\_estimators = 1000$ 、 $max\_depth = 6$  與  $learning\_rate = 0.05$ ，以 RMSE 作為評估指標，並透過梯度提升 (Gradient Boosting) 迭代訓練多棵弱分類樹來提升預測效果。LightGBM 採用  $n\_estimators = 500$ 、 $num\_leaves = 31$  與  $learning\_rate = 0.05$ ，其特點是基於葉節點生長策略 (Leaf-

wise Growth) 以提高效率並降低記憶體消耗。CatBoost 模型則設定  $iterations = 1000$ 、 $depth = 6$  與  $learning\_rate = 0.05$ ，同樣以 RMSE 作為損失函數，並利用有序提升 (Ordered Boosting) 技術減輕資料偏差。

在深度學習方法部分，本研究設計了一個基於殘差結構的多層迴歸模型，底層採用自定義的 Custom ResNet。其架構包含六個全連接殘差模塊，每一模塊由全連接層、均方根正規化 (RMSNorm)、ReLU 激活函數與 Dropout (0.5) 組成，並透過殘差連接強化深層結構的穩定性與可訓練性。輸入首先經 RMSNorm 正規化，隨後依序通過六個殘差模塊，並由兩層全連接層完成降維與映射，最終輸出長度為二的向量，分別對應效價 (Valence) 與喚醒度 (Arousal)。在輸出階段，模型引入正弦函數作為平滑激活，並透過縮放和平移操作將預測值限制於 1 至 9 的範圍，其轉換公式如下：

$$\hat{y} = \sin(x) \times 4 + 5 \quad (3)$$

其中， $x$  為模型線性層的輸出， $\hat{y}$  為最終預測分數，確保結果符合任務的標註需求。

### 3.4 實驗結果

首先，表 1 顯示了不同訓練資料組合的效能表現。本研究的嵌入模型採用 DeepSeek-R1-0528-Qwen3-8B，負責將輸入文本轉換為高維語意向量；迴歸模型則選用支持向量迴歸 (SVR)，用以進行情緒維度的連續數值預測。在僅使用訓練集的情況下，Valence 與 Arousal 的 MAE 分別為 0.528 與 0.914，而 PCC 分別為 0.711 與 0.483。當額外引入部分驗證集樣本時，模型效能些微改善，最終在

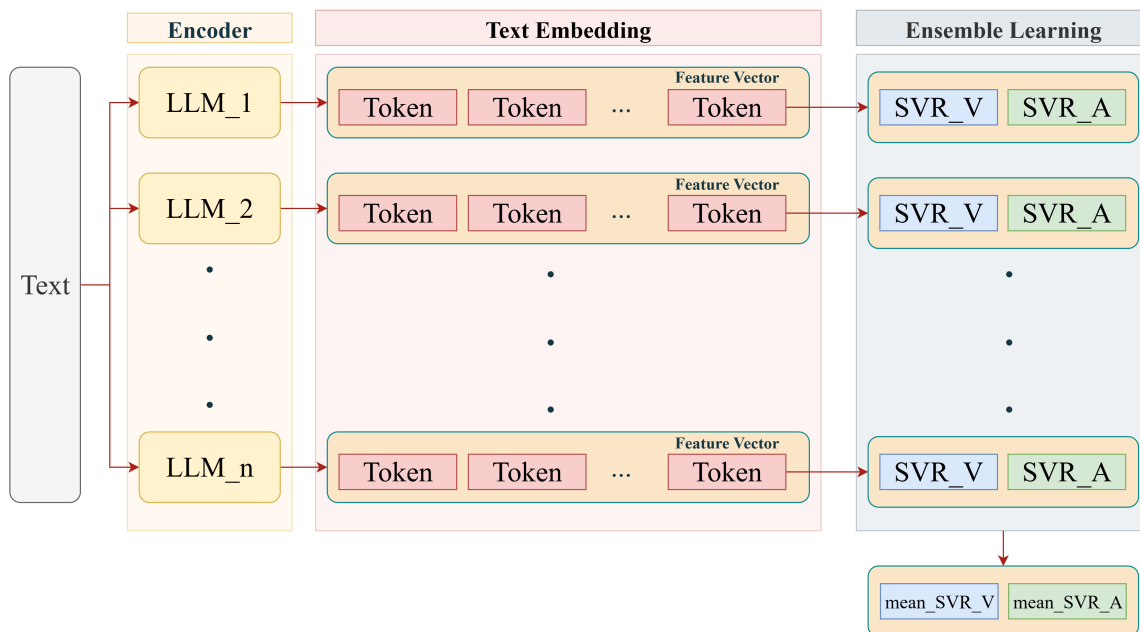


Figure 1: 多編碼器集成學習架構

	MAE_V	MAE_A	PCC_V	PCC_A
Models	0.495	0.802	0.772	0.544
Encoders	0.463	0.759	0.805	0.608

Table 4: 集成式方法

「訓練集 + 驗證集一半」的配置下達到最佳表現，Valence MAE 為 0.524、Arousal MAE 為 0.809，對應的 PCC 分別為 0.754 與 0.538。

進一步比較不同機器學習模型（表 2），本研究在「Training + half validation」訓練策略下，並以 DeepSeek-R1-0528-Qwen3-8B 所生成的嵌入特徵進行實驗。結果顯示，SVR 與 Custom ResNet 的整體表現優於其他梯度提升類模型。SVR 在四項指標上均保持穩定，其中 Valence PCC 為 0.754，Arousal PCC 為 0.538，整體排名第一。Custom ResNet 雖在 Valence PCC (0.761) 與 Arousal PCC (0.542) 上略優，但其 MAE 表現稍遜於 SVR，整體排名第二。這表明 SVR 在小樣本且具跨領域差異的情境中展現出最佳的泛化能力。

在編碼器比較方面（表 3），本研究採用支持向量迴歸 (SVR) 作為迴歸方法，訓練資料使用 Training + validation 進行實驗。結果顯示，DeepSeek-R1-0528-Qwen3-8B 與 multilingual-e5-large-instruct 的整體效能相近，其中前者在 Valence PCC (0.754) 上取得最佳表現，而後者則在 Arousal PCC (0.539) 略優。其他編碼器如

DeepSeek-Prover-V1.5-RL 與 Llama3-TAIDE-LX-8B-Chat-Alpha1 亦展現出一定的穩定性，但整體效能稍低。相較之下，multilingual-e5-large 在四項指標中皆為最弱，顯示專為多語言檢索設計的模型在醫療情境下的適配性有限。整體而言，針對中文醫療文本，DeepSeek 系列編碼器在 Valence 與 Arousal 的預測上展現出更佳的穩定性與適配性。

此外，我們使用集成學習方法進一步提升模型效能，並針對不同層級的集成方式進行比較，其結果如表 4 所示。具體而言，Models 為將表 2（不同迴歸模型）之結果進行集成，即結合 SVR、Custom ResNet 與梯度提升類模型的預測輸出將其取平均值；而 Encoders 則為將表 3（不同編碼器）之結果進行集成，即融合多個 LLM 編碼器所生成的語意嵌入。比較結果顯示，編碼器層級的集成效果顯著優於模型層級集成，特別是在相關性指標上，Valence PCC 達到 0.805，Arousal PCC 為 0.608，均高於模型層級集成的表現 (Valence PCC = 0.772、Arousal PCC = 0.544)。此結果說明，透過融合不同語意表示能更有效捕捉醫療文本中的情緒訊號，對於 DSA 任務

的效能提升尤為關鍵。

最後，圖 1 展示了所提出的多編碼器集成架構。該方法使用不同 LLM 編碼器生成的語意嵌入，在迴歸模型預測 V、A 值後取平均值進行融合，以充分利用多視角的語意特徵。實驗結果證實此架構能進一步提升預測準確性，展現了多編碼器集成學習在醫療語境下進行中文維度情感分析的潛力。

### 3.5 結論

本研究針對 ROCLING-2025 共享任務的中文醫療自我反思文本，系統性地探討了大型語言模型嵌入特徵與多種迴歸方法在維度情感分析任務中的表現，並進行多層次比較與驗證。綜合實驗結果與分析，主要結論如下：

1. 領域適應的重要性：在訓練過程中適度引入目標領域資料（例如採用「Training + half validation」策略），能有效緩解通用語料與醫療文本之間的領域落差，些微提升模型在效價（Valence）與喚醒度（Arousal）預測上的準確度。
2. 迴歸模型比較：在多種迴歸方法中，支持向量迴歸（SVR）與 Custom ResNet 表現最佳。SVR 在少量資料下訓練快速，且超參數需求較低，四項指標整體均衡，展現出穩健的泛化能力；Custom ResNet 則在部分 PCC 指標上略優，但在 MAE 上不及 SVR，整體表現次之。
3. 編碼器效能分析：不同編碼器對情感特徵擷取的能力差異明顯。其中，DeepSeek-R1-0528-Qwen3-8B 在 Valence 預測上達到最佳表現，而 multilingual-e5-large-instruct 則在 Arousal 預測上略佔優勢。整體而言，DeepSeek 系列編碼器在中文醫療文本的情感特徵建模上展現出更高的穩定性與適配性。
4. 集成學習的優勢：相較於單一編碼器或僅在模型層級進行集成，多編碼器的特徵層級集成策略能更有效整合不同模型的優勢，進一步提升模型的泛化能力與穩健性。實驗結果證實，集成架構在 Valence 與 Arousal 的多項指標上均達到最優表現。

綜合而言，本研究證明了將大型語言模型作為特徵提取器，並結合適當的迴歸方法與集成策略，能有效提升中文維度情感分析在醫療場景中的應用價值。此方法不僅在醫療自我反思文本上展現了優異效能，也為發展非侵入式、

基於自然語言的心理健康評估工具提供了技術支撐與實證基礎。

## References

- Rafael A. Calvo and Sidney D'Mello. 2010. [Affect detection: An interdisciplinary review of models, methods, and their applications](#). *IEEE Transactions on Affective Computing*, 1(1):18–37.
- Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 785–794.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Harris Drucker, Chris J.C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. 1997. Support vector regression machines. In *Advances in neural information processing systems*, pages 155–161.
- Daya Guo et al. 2025. [Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning](#).
- Khan Md Hasib, Sami Azam, Asif Karim, Ahmed Al Marouf, F M Javed Mehedi Shamrat, Sidratul Montaha, Kheng Cher Yeo, Mirjam Jonkman, Reda Alhaji, and Jon G. Rokne. 2023. [Mcnlstm: Combining cnn and lstm to classify multi-class text in imbalanced news data](#). *IEEE Access*, 11:93048–93063.
- Jiaxi Hu, Leyuan Qu, Haoxun Li, and Taihao Li. 2025. [Label Semantic-Driven Contrastive Learning for Speech Emotion Recognition](#). In *Inter-speech 2025*, pages 4348–4352.
- James A. Russell Jonathan Posner and Bradley S. Peterson. 2005. [The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology](#). *Development and Psychopathology*, 17(3):715–734.
- Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, pages 3146–3154.

- Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
- Lung-Hao Lee, Jian-Hong Li, and Liang-Chih Yu. 2022. Chinese emobank: Building valence-arousal resources for dimensional sentiment analysis. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(4).
- Lung-Hao Lee, Tzu-Mi Lin, Hsiu-Min Shih, Kuo-Kai Shyu, Anna S. Hsu, and Peih-Ying Lu. 2025. Rocling-2025 shared task: Chinese dimensional sentiment analysis for medical self-reflection texts. In *Proceedings of the 37th Conference on Computational Linguistics and Speech Processing*.
- Dongdong Li, Shengyao Huang, Li Xie, Zhe Wang, and Jiazhen Xu. 2025. Neuron perception inspired eeg emotion recognition with parallel contrastive learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(8):14049–14062.
- Bing Liu. 2012. *Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies*, volume 5. Morgan & Claypool Publishers.
- Nikolaos Malandrakis, Alexandros Potamianos, Elias Iosif, and Shrikanth Narayanan. 2013. Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11):2379–2392.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space.
- Saif M. Mohammad. 2016. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Herbert L. Meiselman, editor, *Emotion Measurement*, pages 201–237. Woodhead Publishing.
- OpenAI. 2023. Gpt-4 technical report. <https://cdn.openai.com/papers/gpt-4.pdf>.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Liudmila Prokhoronkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2017. Catboost: unbiased boosting with categorical features. *arXiv preprint arXiv:1706.09516*.
- Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved semantic representations from tree-structured long short-term memory networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1556–1566, Beijing, China. Association for Computational Linguistics.
- TAIDE. 2025. Llama3-TAIDE-LX-8B-Chat-Alpha1. <https://huggingface.co/taide/Llama3-TAIDE-LX-8B-Chat-Alpha1>. Accessed: 2025-09-01.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention is all you need.
- Krishnapriya Vishnubhotla, Daniela Teodorescu, Mallory J. Feldman, Kristen A. Lindquist, and Saif M. Mohammad. 2024. Emotion granularity from text: An aggregate-level indicator of mental health.
- Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Multilingual e5 text embeddings: A technical report. *arXiv preprint arXiv:2402.05672*.
- Amy B. Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods*, 45(4):1191–1207.
- Aleksandra Wawer. 2024. Few-shot methods for aspect-level sentiment analysis. *Information*, 15(11):664.
- Yunhe Xie, Chengjie Sun, Ziyi Cao, Bingquan Liu, Zhenzhou Ji, Yuanhao Liu, and Lili Shan. 2025. A dual contrastive learning framework for enhanced multimodal conversational emotion recognition. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 4055–4065, Abu Dhabi, UAE. Association for Computational Linguistics.
- Huajian Xin, Z. Z. Ren, Junxiao Song, Zhihong Shao, Wanxia Zhao, Haocheng Wang, Bo Liu, Liyue Zhang, Xuan Lu, Qiushi Du, Wenjun Gao, Qihao Zhu, Dejian Yang, Zhibin Gou, Z. F. Wu, Fuli Luo, and Chong Ruan. 2024. Deepseek-prover-v1.5: Harnessing proof assistant feedback for reinforcement learning and monte-carlo tree search.

- Qi Yan, Zheng Jiang, Zachary Harbin, Preston H Tolbert, and Mark G Davies. 2021. [Exploring the relationship between electronic health records and provider burnout: A systematic review](#). *Journal of the American Medical Informatics Association*, 28(5):1009–1021.
- Kailai Yang, Tianlin Zhang, Hassan Alhuzali, and Sophia Ananiadou. 2023. [Cluster-level contrastive learning for emotion recognition in conversations](#). *IEEE Transactions on Affective Computing*, 14(4):3269–3280.
- Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. [Hierarchical attention networks for document classification](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1480–1489, San Diego, California. Association for Computational Linguistics.
- Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016. [Building Chinese affective resources in valence-arousal dimensions](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 540–545, San Diego, California. Association for Computational Linguistics.
- 林巍. 2022. [多任務維度型情感分析之研究](#). 博士論文, 元智大學, 臺灣博碩士論文知識加值系統.