A Multimodal Learning Approach Based on Chief Complaints for Cognitive Impairment Prediction in Patients with Unipolar and Bipolar Depression

指導教授：胡雅涵、蔡志豐

研究生：吳芮妍

Contents

01

Introduction

Background / Motivation / Research objective

1.1 Background

- Global Distribution
  - Unipolar Depression (UD) (280M cases)
  - Bipolar Depression (BD) (40M cases)
- Economic Burden
  - Current: >$1T annually
  - 2030 Projection: $6T

- UD
  - Persistent depressed mood
  - Anhedonia
  - Neurovegetative symptoms
  - Sleep disturbances
- BD
  - Alternating episodes
  - Energy level fluctuations
  - Impaired judgment
  - Behavioral manifestations

1.2 Motivation

Current Limitations in Mood Disorders

- Limited **predictive modeling**
- **Demographic factors** matter — but are **rarely used** in prediction models
- **Multimodal methods** (Audio + Text + Demographics) work well in **neurodegenerative disorders**

➡️ Yet remain **underused in mood disorder research**

1.3 Research objective

Research Framework Proposal

- **Objective**
  Develop a multimodal framework for CI prediction in **mood disorder patients (UD & BD)** by integrating **demographic**, **audio**, and **text** data.

- **Proposed Framework**
  1. **Multimodal integration** of:
     - Demographic variables (Baseline)
     - Audio features (from patient speech)
     - Text features (from patient speech)
  2. **Feature Combination Evaluation**:
     - Handcrafted features
     - Learned features

02

Literature review

Cognitive impairment in mood disorders
/ Current research on cognitive impairment in mood disorders
/ Multimodal cognitive impairment prediction

2.1 Cognitive impairment in mood disorders

- Major Impaired Domains
  - Executive Function
  - Memory
  - Attention
  - Processing Speed
- Persistence and Progression
  - CI persists **post-remission** and may **worsen** with episode recurrence (UD)
  - CI begins in early episodes and **progresses even in euthymia** (BD)

- Functional Impact
  - ↓ Occupational productivity
  - ↓ Social functioning
  - ↑ Suicide risk

- Clinical Challenges
  - Underutilized assessment, resource constraints, time-intensive testing

→ Need for early detection and targeted cognitive rehabilitation

2.2 Current research on cognitive impairment in mood disorders

2.2.2 Factors associated with cognitive function in mood disorders

- 📊 Age ↑ → Cognitive decline
- ♀♂ Sex differences → Distinct cognitive profiles
- 🎓 Lower education → Poorer performance
- 💔 Unmarried status → Weaker cognitive function
- 🩺 Comorbidities → Increased risk of CI
- 🍺 Alcohol dependence → Long-term cognitive damage
- 🏥 More psychiatric hospitalizations→ Worse cognitive outcomes

⚠ **Critical Gap**

Demographic and clinical data are readily available yet often underused in CI prediction in mood disorders, missing a key opportunity for early detection and intervention.

2.3 Multimodal cognitive impairment prediction

Current Limitations

Predominant focus on neurodegenerative disorders with limited exploration in mood disorders

Validation Requirements

Untested efficacy of multimodal prediction frameworks in mood disorder populations, with particularly limited research integrating demographic data alongside audio and/or text modalities

Development Needs

Mood disorder-specific feature identification and optimal multimodal integration strategies

03

Research method

Data sources and study sample
/Demographic variable
/Feature extraction/Classification methods

3.1 Data sources and study sample

- **Wisconsin Card Sorting Test (WCST) Description**
  - 128 response cards
  - Three dimensions: color, form, number
  - Match cards based on undisclosed principle
  - Binary feedback system

- **Target Variables**
  - WCST4: Total Errors
  - WCST6: Perseverative Responses
  - WCST8: Perseverative Errors
  - WCST12: Categories Completed

Threshold: Normal ($T \geq 40$) / Impaired ($T < 40$)

3.2 Demographic variable

Table Demographic variable definition

Variable	Definition
Sex	Biological sex of the participant
Years of Education	Total years of formal education received by the participant
Education Level	The highest level of education attained by the participant
Age	Chronological age of the participant (years)
Marriage	Marital status of the participant
Residential Status	Residential status, indicating whether the participant lives alone or with others
Occupation	Employment status of the participant
Suicide History	History of suicidal behavior or ideation
Psychiatric Hospitalization History	History of psychiatric hospitalization
Alcohol Consumption History	History of alcohol consumption
Cerebrovascular Accident	History of cerebrovascular accident (stroke)
Intracerebral Hemorrhage	History of intracerebral hemorrhage
Thyroid	History of thyroid dysfunction
Diabetes Mellitus	History of diabetes mellitus
Hypertension	History of hypertension.
Hyperlipidemia	History of hyperlipidemia.
Epilepsy	History of epilepsy

3.3 Feature extraction

Audio Features

Handcrafted Features
- Prosodic Features (7)
- eGeMAPSv01a Features (88)

Learned Features
- VGGish (128)
- Wav2vec 2.0 (1024)
- HuBERT (1024)

Text Features

Handcrafted Features
- TF-IDF (99)
- LIWC (91)

Learned Features
- TAIDE (4096)
- ERNIE-Health (768)
- RoBERTa (768)

3.4 Classification methods

- SVM
  - Optimal hyperplane construction for class separation
- KNN
  - Similarity-based classification using nearest neighbors
- LR
  - Probability-based binary classification
- RF
  - Ensemble method with multiple decision trees
- XGBoost
  - Advanced gradient boosting with tree pruning

04

Results

Baseline characteristics and cognitive comparisons
/Baseline models
/Experimental results

4.1 Baseline characteristics and cognitive comparisons

Significant Variables Between Normal and Impaired Groups Across WCST Measures

WCST-4 (Total Errors)	WCST-6 (Perseverative Responses)
• Years of Education • Education Level • Age • Psychiatric Hospitalization History	• Psychiatric Hospitalization History
WCST-8 (Perseverative Errors)	**WCST-12 (Categories Completed)**
• Years of education • Education Level • Psychiatric Hospitalization History	• Years of Education • Education Level • Age • Marriage • Psychiatric Hospitalization History • Hypertension • Hyperlipidemia

4.2 Baseline models

- **Key Findings Summary**
  - XGB algorithm demonstrates superior balanced performance across all WCST metrics
  - XGB exhibits significant accuracy improvement compared to KNN
  - XGB achieves substantially higher precision metrics than alternative algorithms
  - While KNN shows marginally higher AUC values in certain cases, differences are minimal (≤0.01)
  - XGB's F1 scores exceed KNN indicating more robust classification performance

4.3 Experimental results

WCST-4 (Total Errors)	WCST-6 (Perseverative Responses)
• PF + Dem • H + Dem • W + Dem	• H + Dem • EG + Dem • V + Dem
WCST-8 (Perseverative Errors)	**WCST-12 (Categories Completed)**
• W + Dem • H + Dem • EG + Dem	• PF + Dem • H + Dem • W + Dem

Note(s): PF = Prosodic features, EG = eGeMAPSv01a features, V = VGGish, H = HuBERT, W = Wav2vec2.0, AUC = Area under the curve, SVM = Support vector machine, KNN = K-nearest neighbors, LR = Logistic regression, RF = Random forest, XGB = Extreme gradient boosting.

4.3 Experimental results

WCST-4 (Total Errors)	WCST-6 (Perseverative Responses)
• LIWC + Dem • E + Dem • TF-IDF + Dem	• E + Dem • TA + Dem • LIWC + Dem
WCST-8 (Perseverative Errors)	**WCST-12 (Categories Completed)**
• E + Dem • LIWC + Dem • TA + Dem	• LIWC + Dem • E + Dem • TF-IDF + Dem

Note(s): TF-IDF = Term frequency-inverse document frequency, LIWC = Linguistic inquiry and word count, TA = TAIDE, E = ERNIE-Health, R = RoBERTa, AUC = Area under the curve, KNN = K-nearest neighbors, LR = Logistic regression, RF = Random forest, XGB = Extreme gradient boosting.

4.3 Experimental results

WCST-4 (Total Errors)	WCST-6 (Perseverative Responses)
• PF + LIWC + Dem • PF + TF-IDF + Dem • H + E + Dem	• V + E + Dem • V + TA + Dem • H + TA + Dem
WCST-8 (Perseverative Errors)	**WCST-12 (Categories Completed)**
• W + E + Dem • H + E + Dem • H + TA + Dem	• PF + LIWC + Dem • PF + TF-IDF + Dem • W + E + Dem

Note(s): PF = Prosodic features, V = VGGish, H = HuBERT, W = Wav2vec2.0, TF-IDF = Term frequency-inverse document frequency, LIWC = Linguistic inquiry and word count, TA = TAIDE, E = ERNIE-Health, AUC = Area under the curve, SVM = Support vector machine, KNN = K-nearest neighbors, RF = Random forest, XGB = Extreme gradient boosting.

4.3 Experimental results

- **Overall:** Superior to demographic variables alone across all WCST measures
- **Modality Comparison**:
  - Outperforms audio-based: WCST-6, 8, 12
  - Outperforms text-based: WCST-4, 12
  - Text-based superior for perseverative measures (WCST-6, 8)
- **Optimal Feature Combinations**
  - WCST-4: Demographic + Handcrafted audio
  - WCST-6/8: Demographic + Learned text
  - WCST-12: Demographic + Handcrafted audio + Handcrafted text

05

Discussion

Theoretical implications
/Clinical implications

5 Discussion

- **Significant Clinical and Demographic Correlates of CI**
  - Psychiatric hospitalization history: significant factor across all WCST measures (φ = 0.21-0.31)
  - Education: lower in Impaired groups for WCST-4, WCST-8, and WCST-12
  - Marital status: contradictory findings - married individuals overrepresented in Impaired group (WCST-12)
- **Modality-Specific Advantages for Different WCST Measures**
  - Audio-based approach : superior for predicting total errors (WCST-4)
  - Text-based approach : optimal for perseverative responses (WCST-6) and errors (WCST-8)
  - Multimodal approach: significant improvement for categories completed (WCST-12)
- **Optimal Feature Extraction Methods Vary by Cognitive Domain**
  - Total errors (WCST-4): handcrafted prosodic features performed best
  - Perseverative tendencies (WCST-6, WCST-8): ERNIE-Health learned features excelled
  - Categories completed (WCST-12): integration of prosodic features with LIWC features optimal

06

Limitations and future research

6 Limitations and future research

Small dataset	• Incorporate larger, more diverse datasets
Cross-sectional design	• Implement longitudinal studies
Data imbalance	• Explore alternative techniques
Combined disorder model	• Develop disorder-specific models
Limited assessment methodologies	• Incorporate digital physiological features and visuomotor processing with kinematic movement analysis

Thanks

Your comments and feedback are highly appreciated.