Multi-modality Hierarchical Recall based on GBDTs for Bipolar Disorder Classification

Xiaofen Xing, Bolun Cai, Yinhu Zhao,
Shuzhen Li, Zhiwei He, Weiquan Fan

South China University of Technology

Outline

- Introduction
- Feature Extraction
- Multi-modal Hierarchical Recall Framework
- Experiments
- Further Works
- Conclusions

Introduction — What is Bipolar Disorder?

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes

Article Talk

Bipolar disorder

From Wikipedia, the free encyclopedia
(Redirected from Bipolar Disorder)

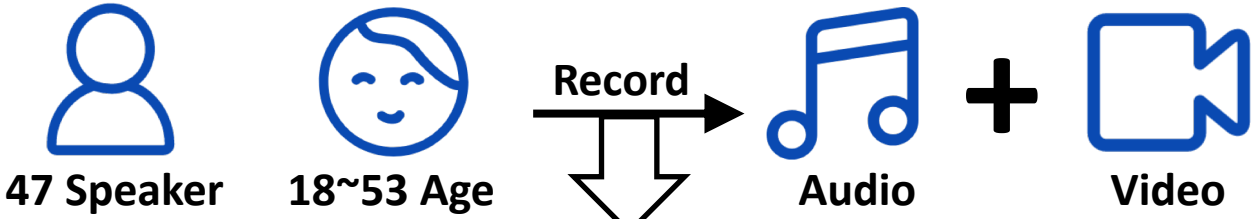
"Manic depression" redirects here. For other uses, see Manic depression (disambiguation).
"Bipolar disorders" redirects here. For the medical journal, see Bipolar Disorders (journal).

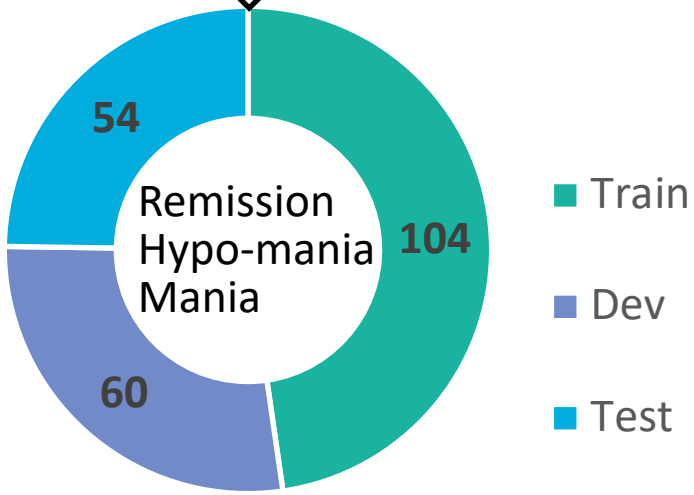
Bipolar disorder, previously known as **manic depression**, is a mental disorder that causes periods of depression and periods of abnormally elevated mood.[3][4][6] The elevated mood is significant and is known as mania or hypomania, depending on its severity, or whether symptoms of psychosis are present.[5] During mania, an individual behaves or feels abnormally energetic, happy, or irritable.[3] Individuals often make poorly thought out decisions with little regard to the consequences.[4] The need for sleep is usually reduced during manic phases.[4] During periods of depression, there may be crying, a negative outlook on life, and poor eye contact with others.[3] The risk of suicide among those with the illness is high at greater than 6 percent over 20 years, while self-harm occurs in 30–40 percent.[3] Other mental health issues such as anxiety disorders and substance use disorder are commonly associated with bipolar disorder.[3]

- **Bipolar disorder** (BD, also called as manic depression) is a **mental disorder** that causes periods of **depression** and abnormally **elevated mood**.

Introduction — AVEC Bipolar Disorder Dataset

The **AVEC Bipolar Disorder Sub-Challenge Dataset** is part of the Turkish Audio-Visual Bipolar Disorder Corpus.





- BD Dataset includes **audio** and **video** recordings of structured interviews performed by **47 speaking** subjects **aged 18-53**.
- There are **104** recordings in training set, **60** recordings in development set, and **54** recordings in test set. The label is divided into **remission**/**hypo-mania**/**mania** according to Young Mania Rating Scale (YMRS).

Introduction—Bipolar Disorder Challenge

- **Fuzzy Label**

  Hard division based YMRS score

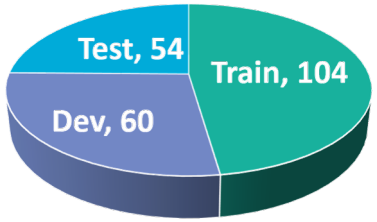
- **Small Sample (Over-fitting)**

  Only 104 recordings in the training set

- **Multiple Modal**

  Including audio and video recordings

$1.Remission: Y_t \leq 7$
$2.Hypo-mania: 7 < Y_t < 20$
$3.Mania: Y_t \geq 20$



Fuzzy Label | **Small Sample** | **Multiple Modal**

Multi-modal Hierarchical Recall Framework

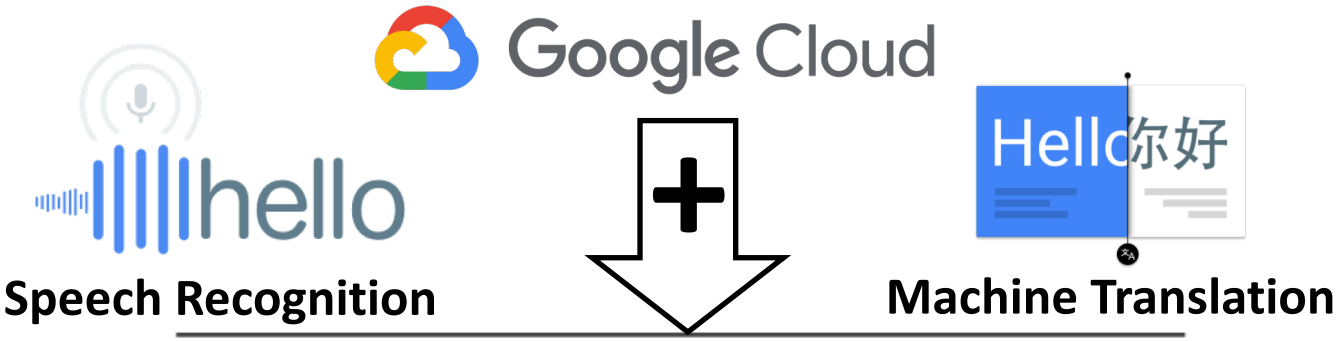
Feature Extraction—Topic Modeling

South China University of Technology

Fine-grained features extracted by topic models will effectively improve the performance.

- Detailed information of each topic is retained.
- Different topic can be described by different features.

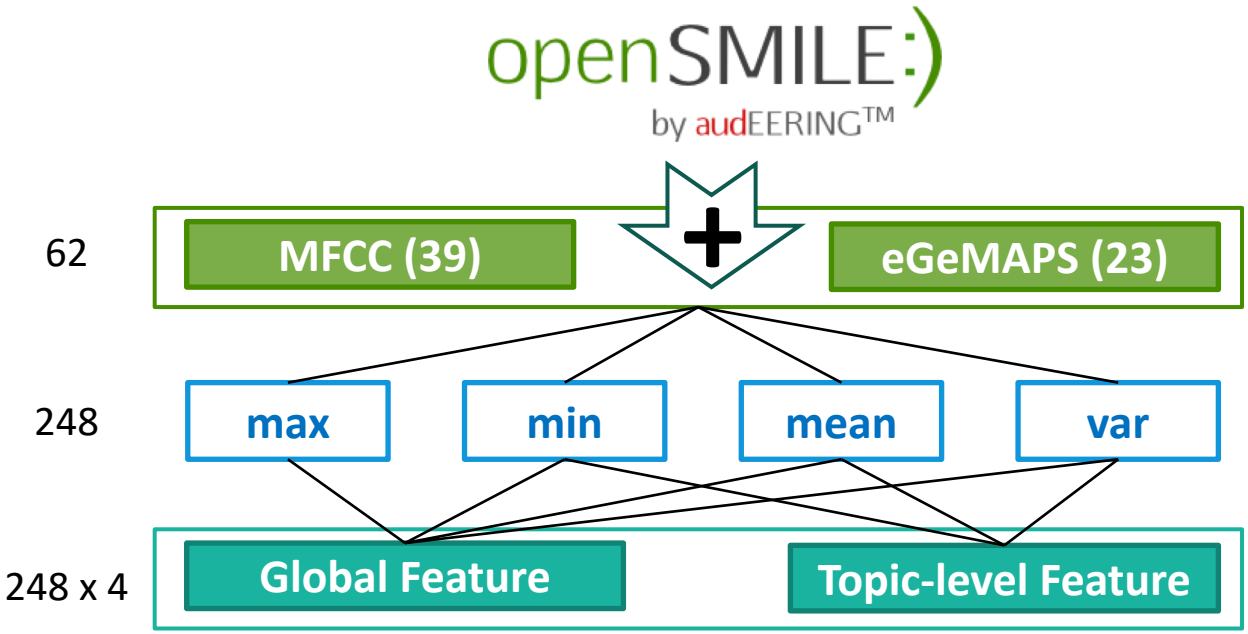
We use **Google Cloud** Platform to transform *Turkish audio* into *English text* with *time-stamps*.



Ind.	Topic Abbr.	Required Task
1	negative_task	Describe why you come here Depict Van Gogh's *Depression* Describe the worst memory
2	neutral_task	Count 1-30 Count 1-30 again (often faster)
3	positive_task	Depict Dengel's *Home Sweet Home* Describe the best memory

Feature Extraction—Audio

Audio features are extracted with openSMILE, including MFCCs (39), and eGeMAPS (23), which are common features for audio analysis.



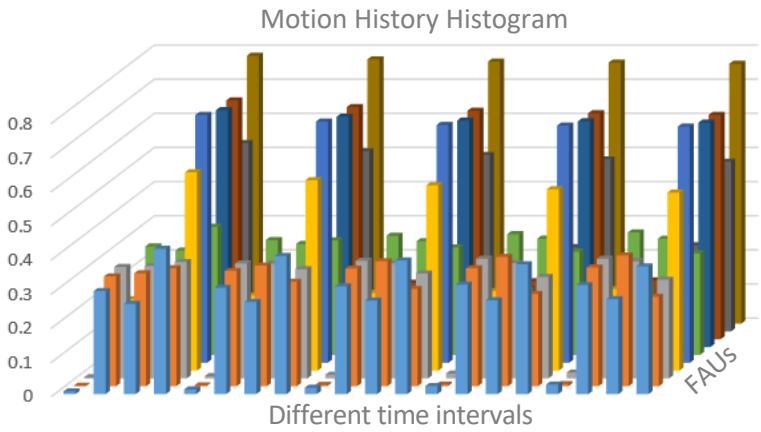
$$
\begin{cases}
f_{smile}^{\max} = \max\limits_{t\in\tau} f_{audio}(t) \\
f_{smile}^{\min} = \min\limits_{t\in\tau} f_{audio}(t) \\
f_{smile}^{mean} = \frac{1}{\tau}\sum\limits_{t\in\tau} f_{audio}(t) \\
f_{smile}^{var} = \sqrt{\frac{1}{\tau}\sum\limits_{t\in\tau}\left(f_{smile}(t) - f_{smile}^{mean}\right)^2}
\end{cases}
$$

- We further concatenate the **statistics features** (62 x 4=248), including max, min, mean, and var, for each topic.
- To describe **global** and **local** information, we respectively extract the **global feature** (248) and **topic-level feature** (248 x 3).

Feature Extraction—Video

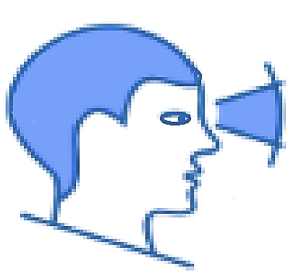
Motivated by Young Mania Rating Scale (YMRS), two kinds visual features (FAUs, eyesight) are related to bipolar disorder.

➢ Facial Action Units (FAUs)



- We extract the FAU features (17x5x10=850) with **5 time intervals** $M_k \in \{10, 20, 30, 40, 50\}$ and **10 equally bins** $R_b \in \{-5, ..., 3, 4, 5\}$
- We apply **16 statistics** to reduce the feature dimension and generate the FAU statistic feature (17×16=272)
- **FAU Feature = (850 + 272) x 3 topic = 3366**

➢ Eyesight

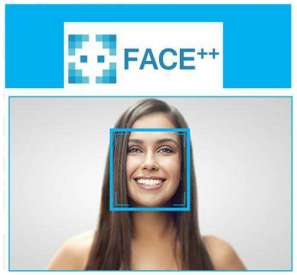


- We extract **7 statistic** features for eyesight, such as min, max, mean, var, std, etc.
- For **left/right eyes** (7x2) and **global head** (3) movement, there are 7x2+3=17 statistic features for each topic.
- **Eyesight Feature = 17 x 3 topic = 51**

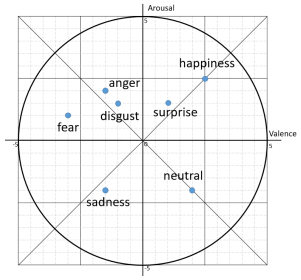
Feature Extraction—Others

In addition, we try to extract the other feature (video, audio and text) related to YMRS, but they did not be used in the final model.

Emotion

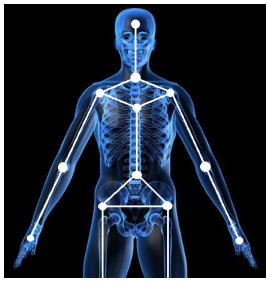


Face++



Valence-Arousal Space

Body Movement



Body movement

Motion History Histogram

Text Feature

Using the suite of Linguistic Analysis Tools (SALAT), text analysis was performed automatically on the transcripts of BD interview.

SenticNet
Helping machines to learn, leverage, love.

SiNLP

EmoLex
Word-Emotion Lexicon

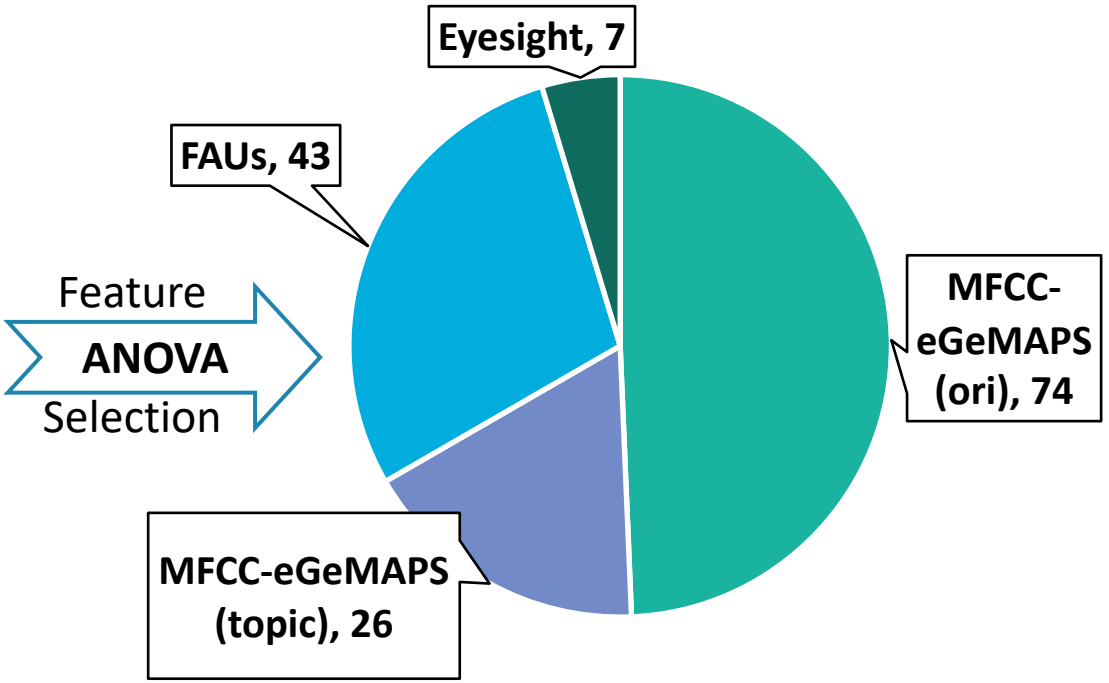
ANEW

LASSWELL

Feature Extraction—Selection

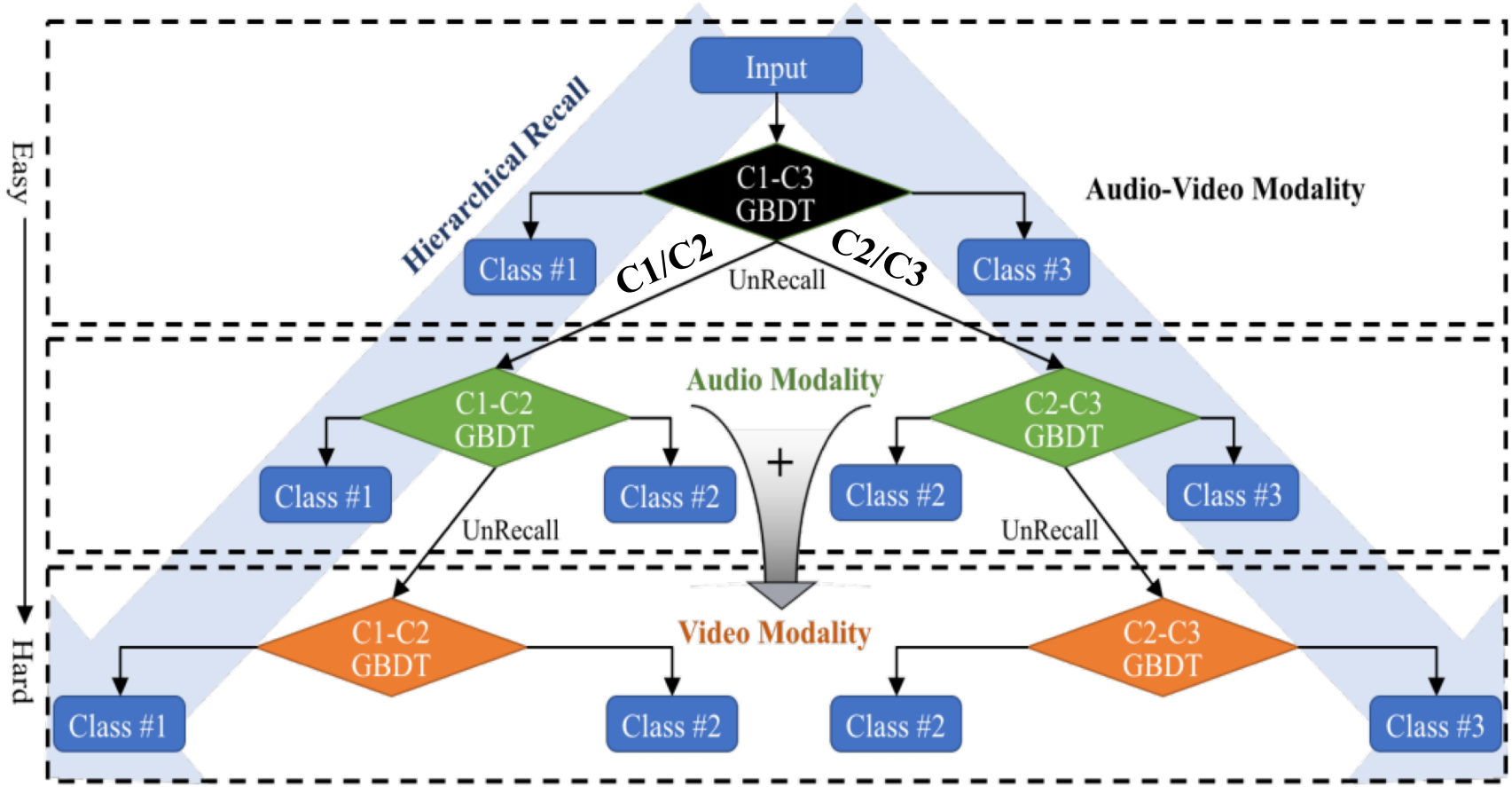
The high-dimensional feature is too redundant to train a classifier with robust performance. To avoid overfitting, we apply Analysis of Variance (ANOVA) method for feature selection.

Feature name	Dimension
MFCC-eGeMAPS-original	248
MFCC-eGeMAPS-3topics	744
Timing	10
AUs	3366
Emotion	183
Eyesight	51
Body movement	7
SiNLP	42
SenticNet	90
ANEW	96
EmoLex	120
Lasswell	438
Sum	5395



- According to feature selection, the official features (MFCC, eGeMAPS, FAUs, eyesight) are the very effective features.

Multi-modal Hierarchical Recall

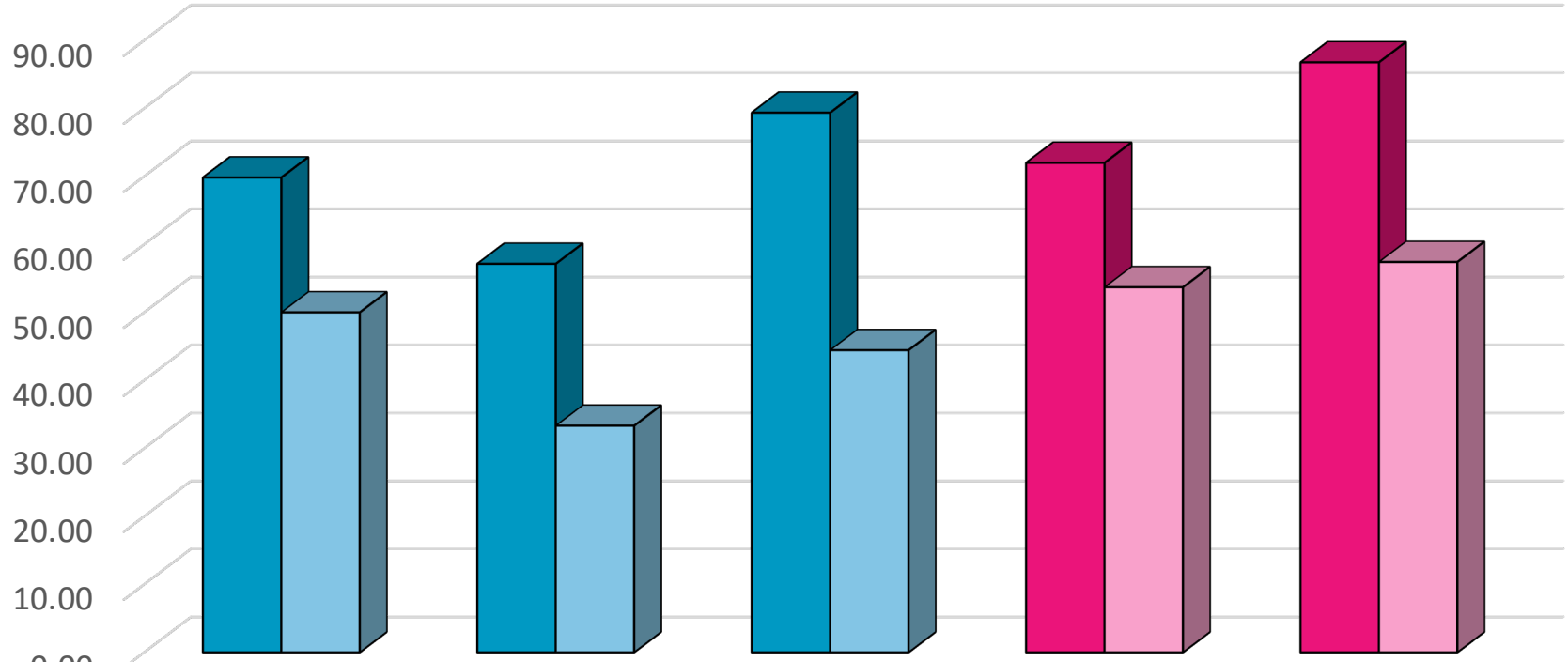


Fuzzy Label: C2 is the margin label and hard to be classified. We use a C1/C2-C2/C3 GBDT for coarse classification.

Small Sample: A few training recordings lead to over-fitting. We use hierarchical GBDTs to recall confidence samples step by step.

Multiple Modal: We use hierarchical modal (audio in the 2-nd layer and video in the 3-th layer) to combine different modal feature.

Experiments — Result

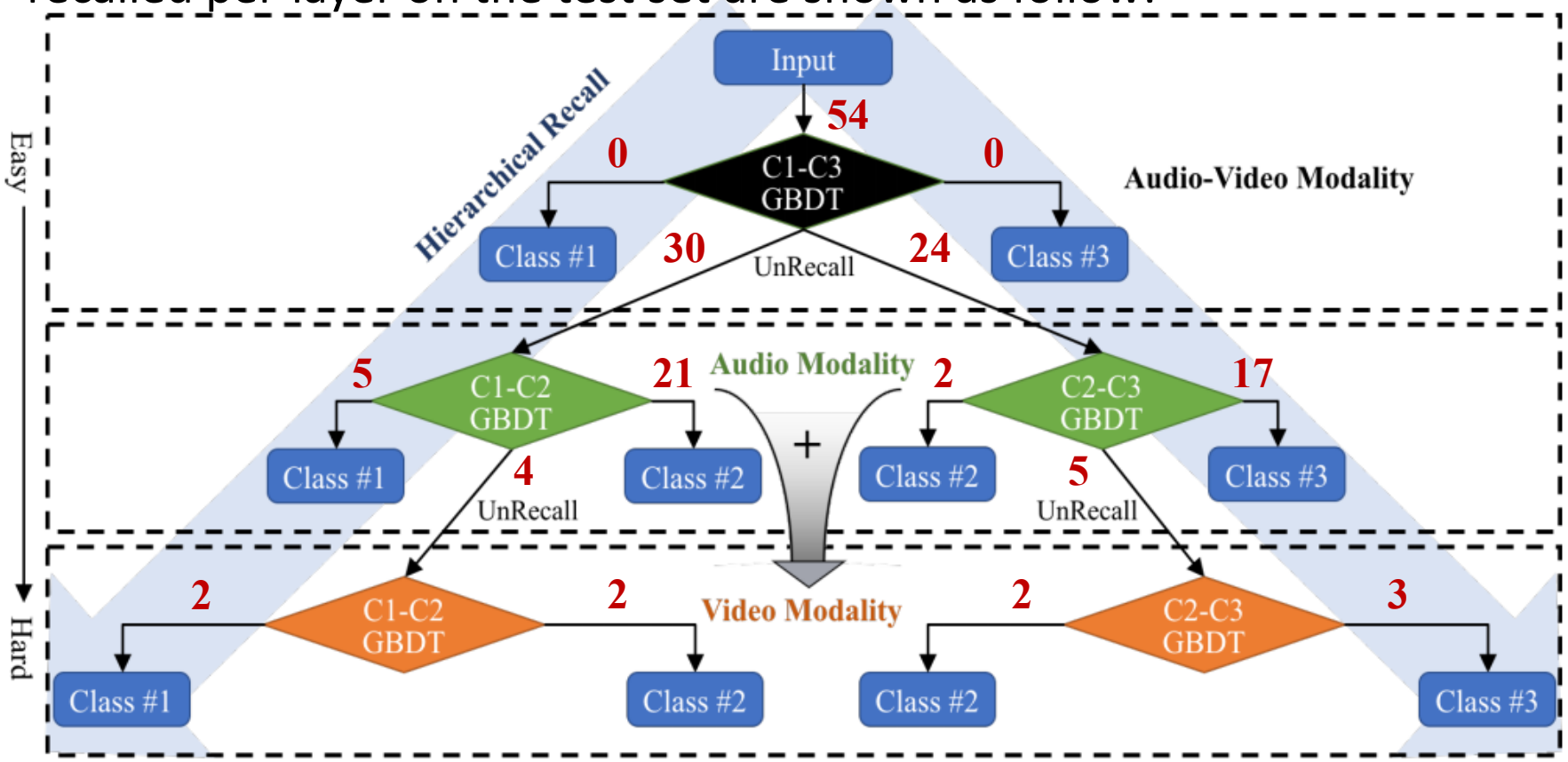


	Baseline (A)	Baseline (V)	Baseline (A+V)	w/o Hierarchy	Ours
Dev	69.84	57.14	79.37	72.00	86.77
Test	50.00	33.33	44.44	53.70	57.41

- Our framework achieves competitive performance by **hierarchical recall**. The result in the dev set is better than the others, and the result in the test set is still clearly better than all baseline. The results show that the proposed framework has **good generalization**.

Experiments — Hierarchical Recall

The parameters of the framework and the number of samples recalled per-layer on the test set are shown as follow:



Layer	Modality	Model	Learning rate	L2 regularization parameter	Threshold
Layer1	Audio+Video	C1-C3	0.03	0.2	0.95
Layer2	Audio	C1-C2	0.13	0.1	0.55
		C2-C3	0.09	0.5	0.55
Layer3	Video	C1-C2	0.19	0.1	\
		C2-C3	0.01	0.1	\

Further Works

South China University of Technology

How can improve the performance further ?

3-topic model

Ind.	Topic Abbr.	Required Task
1	negative_task	Describe why you come here Depict Van Gogh's *Depression* Describe the worst memory
2	neutral_task	Count 1-30 Count 1-30 again (often faster)
3	positive_task	Depict Dengel's *Home Sweet Home* Describe the best memory

Google Cloud + Handcraft Check

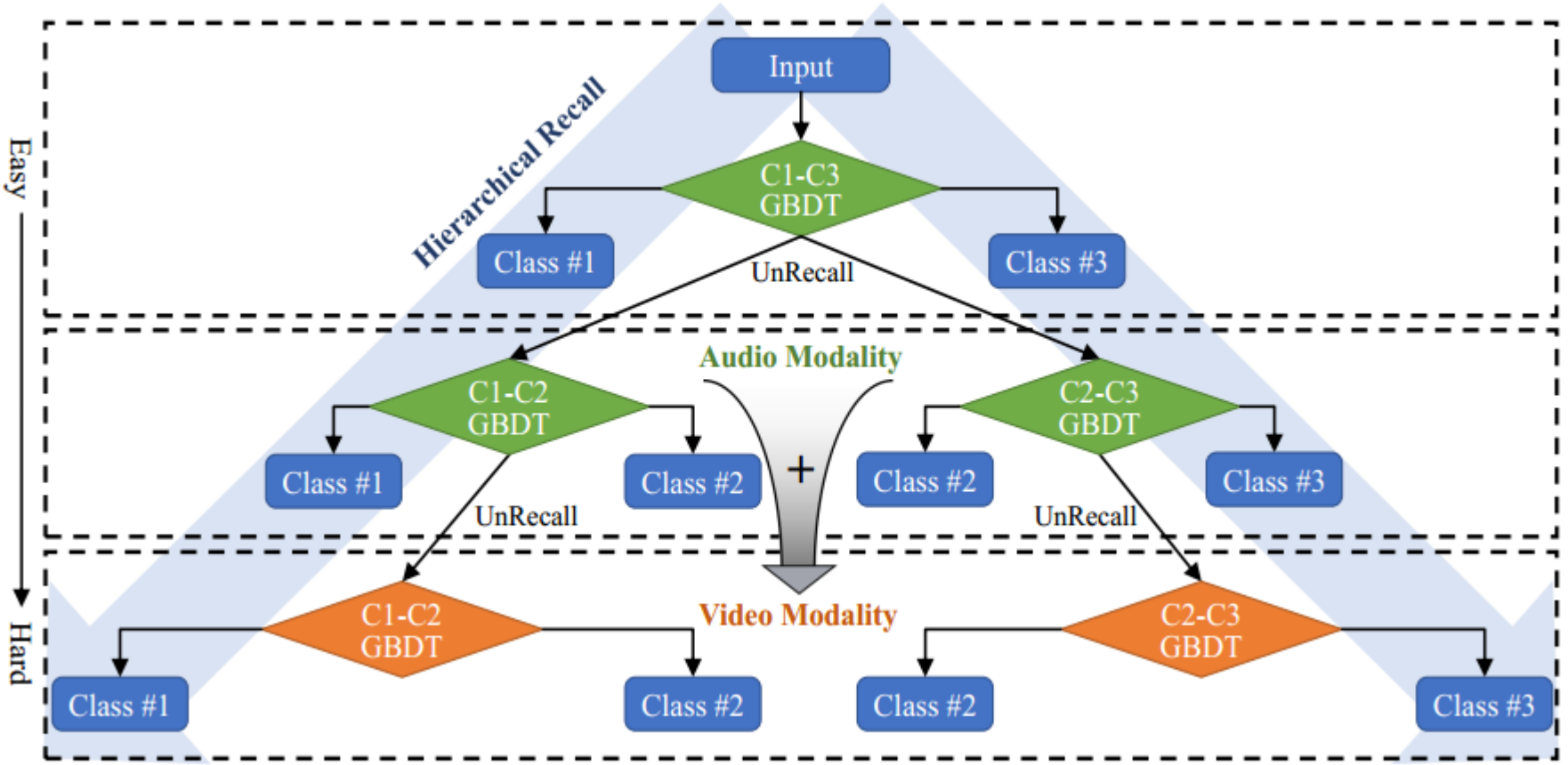
7-topic model

Ind.	Topic Abbr.	Required Task
1	why_come	Describe why you come here
2	man_pic	Depict Van Gogh's *Depression*
3	worst_memory	Describe the worst memory
4	count1	Count 1-30
5	count2	Count 1-30 again (often faster)
6	family_pic	Depict Dengel's *Home Sweet Home*
7	best_memory	Describe the best memory

➢ 7-topic model will bring an improvement compared to 3-topic model

Further Works

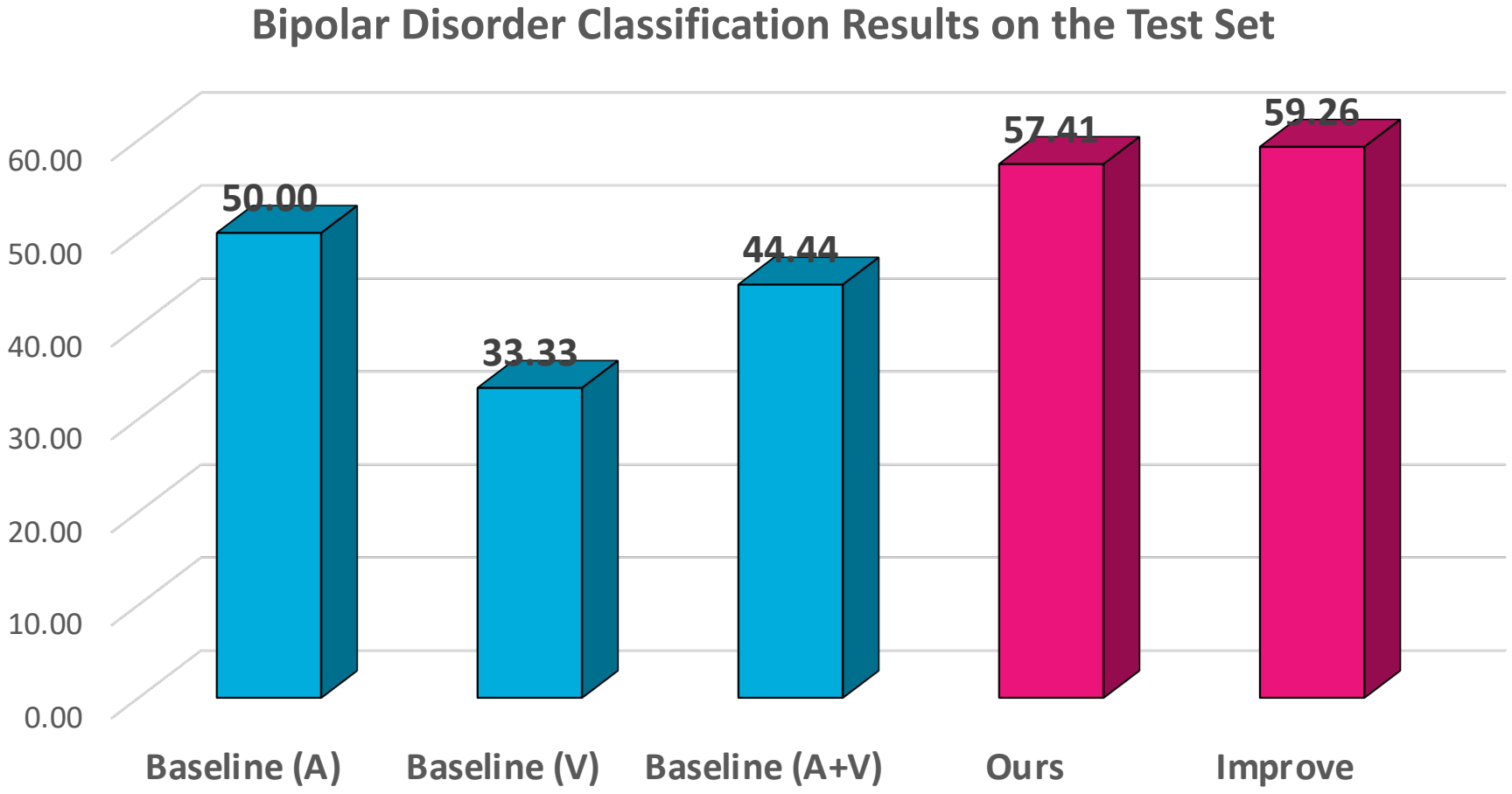
We re-design the multi-modal hierarchical recall framework as follow:



Layer	Modality	Model	Learning rate	L2 regularization parameter	Number of trees	Threshold
Layer1	Audio	C1-C3	0.2	0.1	38	0.85
Layer2	Audio	C1-C2	0.15	0.6	69	0.75
		C2-C3	0.3	0.2	34	0.75
Layer3	Video	C1-C2	0.02	0.1	10	\
		C2-C3	0.08	0.1	4	\

Further Works

South China University of Technology



- Based on 7-topic model, the improved framework can achieve **59.26%** accuracy on the test set, which is higher than the 3-toptic result (57.41%) reported in the paper.

The code can be found at: https://github.com/caibolun/AVEC-BDS2018

Further Works — Transcripts of AVEC BDS

South China University of Technology

Coarse time-stamps and transcripts are auto generated by the **Google Cloud** Platform, and accurate time-stamps and transcripts of each topic are obtained after **handcraft checking**.

Google Cloud + Handcraft Check

start_time	end_time	topic	transcript	confidence
00:04	00:14	1	press the Spacebar key. I did not have information to come here	0.8263
00:15	00:34	1	my friend took me for cup of coffee, they brought coffee and cake and they brought me here. I did not know why I brought it here they said goodness for you for your goodness, nothing else	0.8776
00:36	00:40	0	Will you tell what makes you think what makes you feel.	0.8101
00:40	00:46	2	A man	
00:48	00:52	2	crying or thinking of something Something else	0.8446
00:56	00:58	3	a bad memory	0.7097
00:59	01:08	3	this is bad memory. They brought me here.	0.9058
01:08	01:23	4	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30.	
01:23	01:28	0	fast	
01:28	01:41	5	1 2 3 4 5 6 7 8 9 10 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30	
01:47	01:59	6	a poor family with a buckle There are 5 children or 6 children alive and they have animals	0.7740
02:05	02:07	7	there is so much to join us	0.7193
02:08	02:21	7	but one I'm always happy I live in my home laughing I invite my friends to drink coffee so everyday for me	0.8680

- We hope these tagging can help the other researchers to validate the effective of NLP technology for bipolar disorder classification.

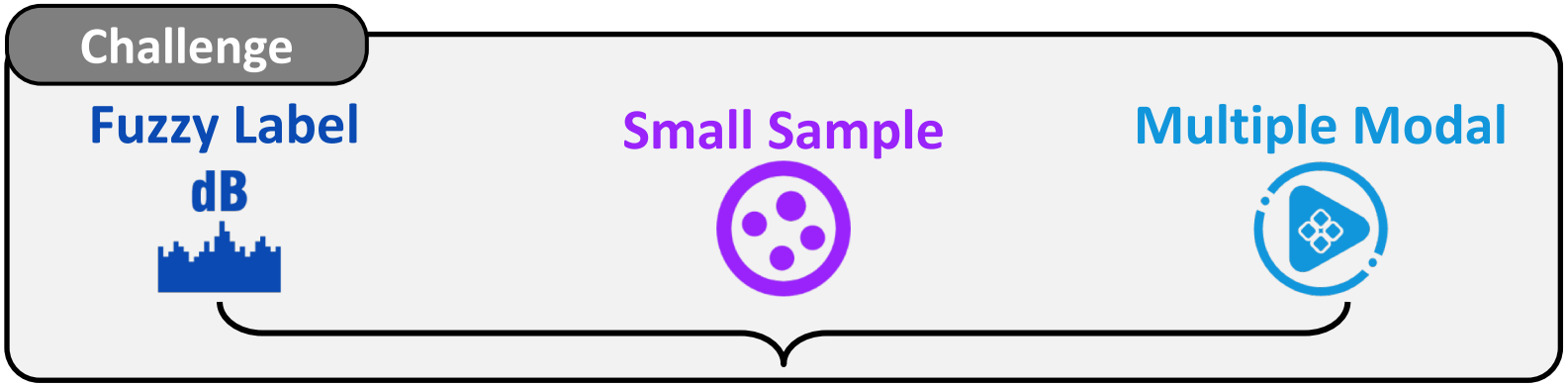
https://github.com/caibolun/AVEC-BDS2018/tree/master/data/translation_check

Conclusions

ACM multimedia Korea mm

AVEC 2018

Bipolar Disorder and Cross-cultural Affect



Multi-modal Hierarchical Recall Framework

Simple: Only official features and GBDTs are used for bipolar disorder classification.

Effective: On the test set, 3-topic model achieves 57.41%, and 7-topic model achieves 59.26%.

Share: We label accurate time-stamps and transcripts to help the others validating NLP to bipolar disorder classification.

https://github.com/caibolun/AVEC-BDS2018/

Thank You

Q & A

caibolun@gmail.com

https://caibolun.github.io/