Detecting Depression with Audio/Text Sequence Modeling of Interviews

Tuka Alhanai, Mohammad Ghassemi, and James Glass
Massachusetts Institute of Technology
Cambridge MA USA

2nd September, Interspeech 2018

Email: tuka@mit.edu Website: talhanai.com Github: github.com/talhanai

How have you been feeling lately?

Healthy: “I’ve been feeling good lately”

Depressed: “*sigh* stressed [um] lately I’ve been really sad and I don’t know why”

Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

Hypothesis: Depression can be modeled using sequences of a spoken interaction.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

Current Solutions

1. **Clinical:**
   Interview/questionnaire

How often during the past 2 weeks were you bothered by...	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things	0	1	2	3
2. Feeling down, depressed, or hopeless	0	1	2	3

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

Current Solutions

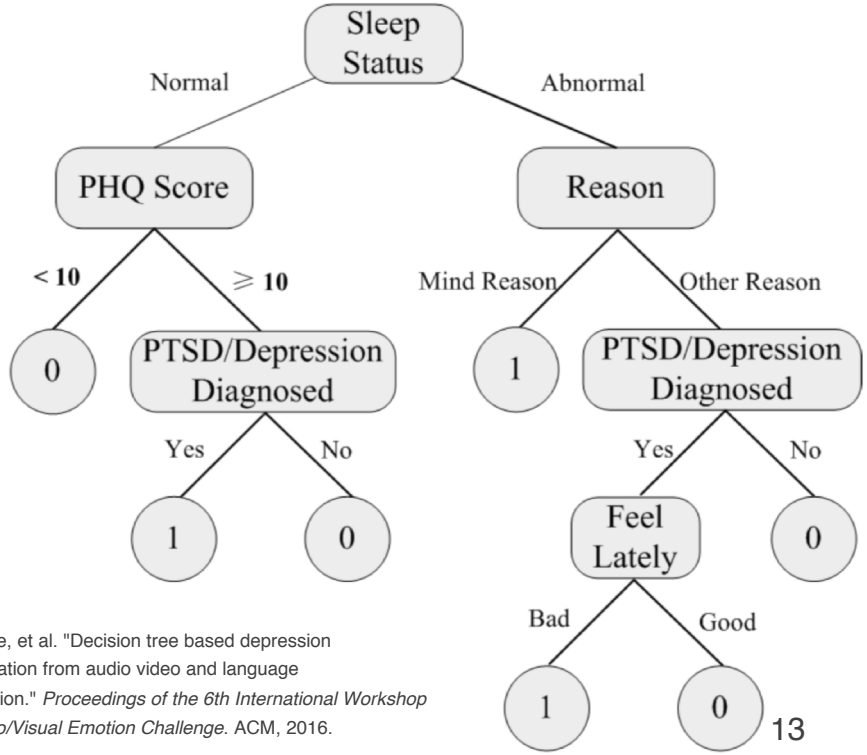
1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

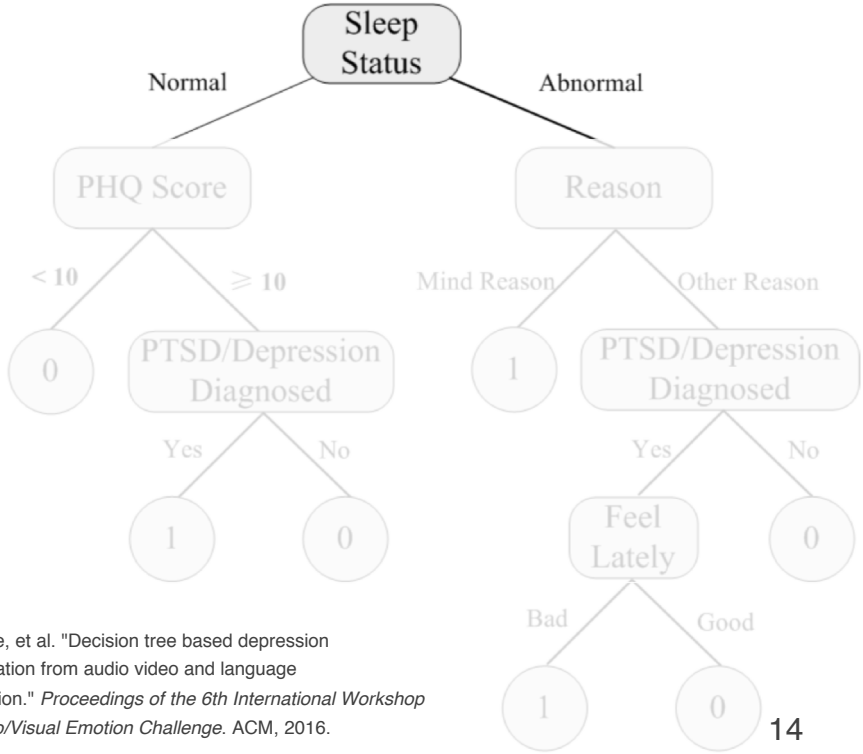


Yang, Le, et al. "Decision tree based depression classification from audio video and language information." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.* ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

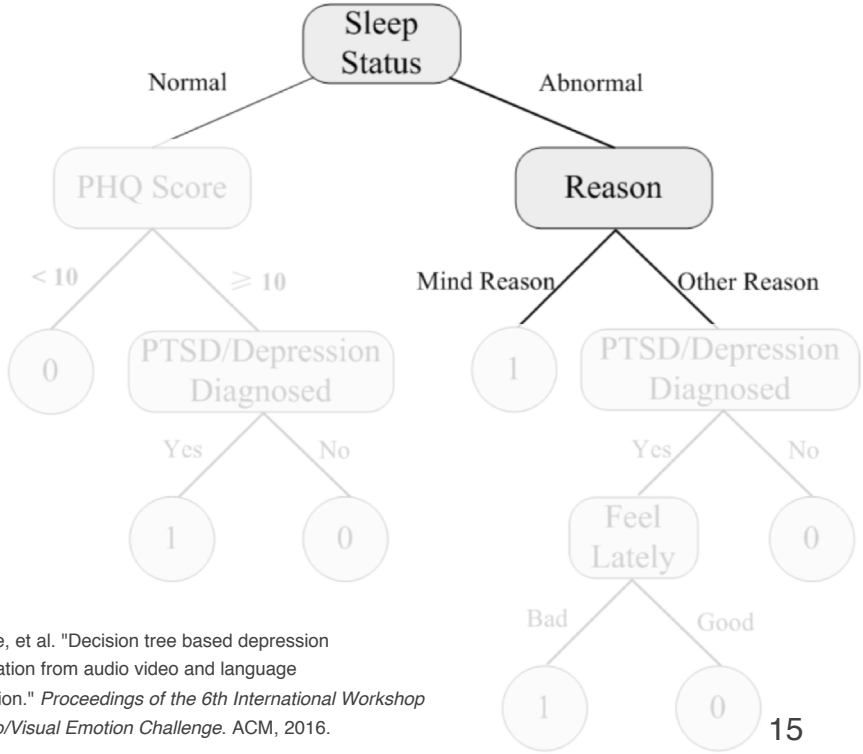


Yang, Le, et al. "Decision tree based depression classification from audio video and language information." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.* ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

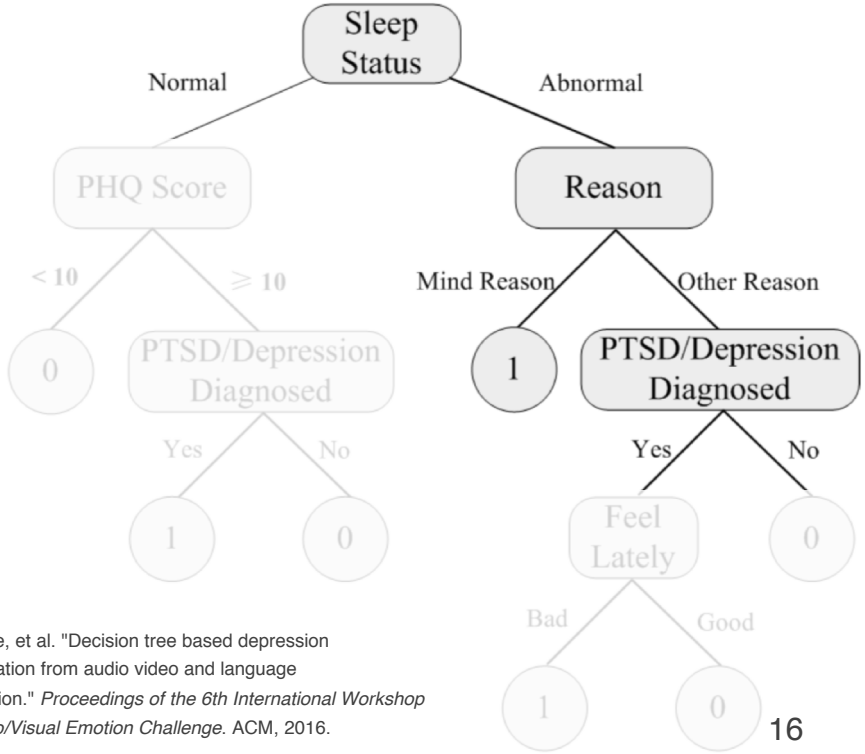


Yang, Le, et al. "Decision tree based depression classification from audio video and language information." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.* ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

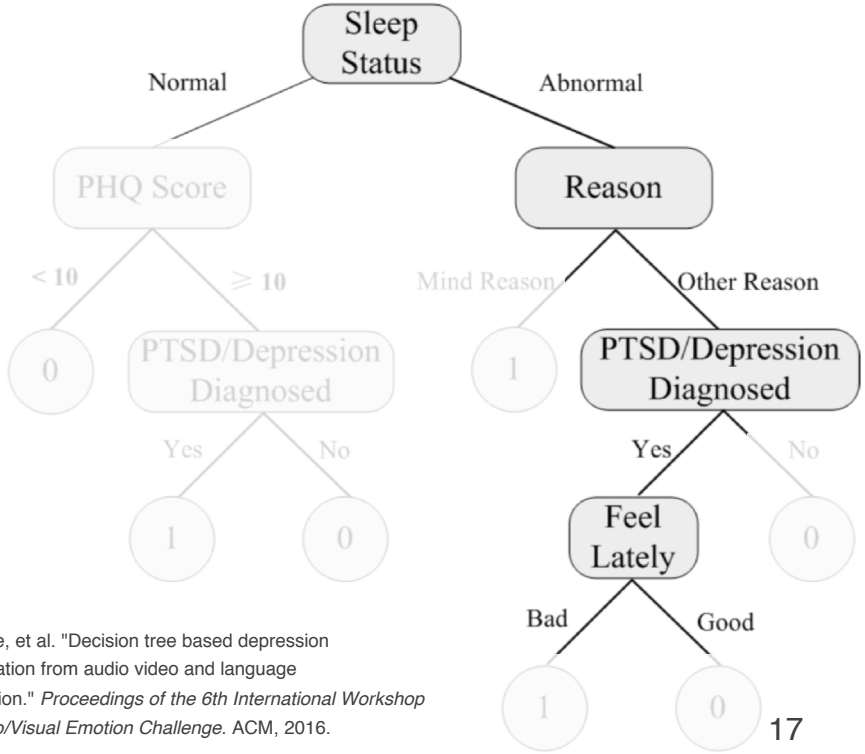


Yang, Le, et al. "Decision tree based depression classification from audio video and language information." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.* ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

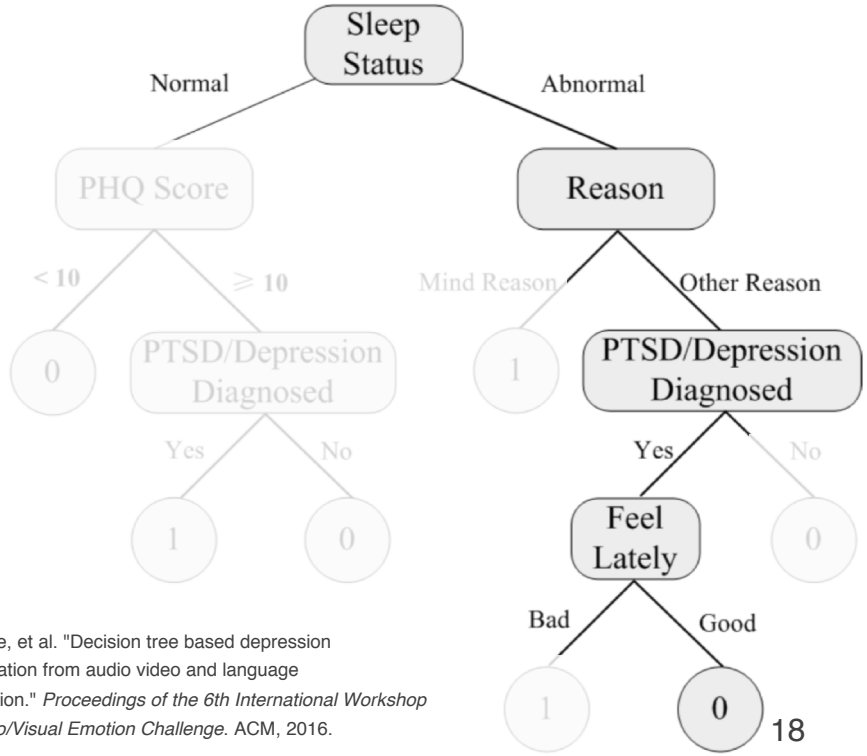


Yang, Le, et al. "Decision tree based depression classification from audio video and language information." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.* ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering



Yang, Le, et al. "Decision tree based depression classification from audio video and language information." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge.* ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

keyword: “therapy”
“sad”
“down”

Williamson, James R., et al. "Detecting depression using vocal, facial and semantic communication cues." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2016.

Current Solutions

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

keyword: “therapy”
“sad”
“down”

Williamson, James R., et al. "Detecting depression using vocal, facial and semantic communication cues." *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, 2016.

[Topic 1 features,
Topic 2 features,
Topic 3 features]

Gong, Yuan, and Christian Poellabauer. "Topic Modeling Based Multi-modal Depression Detection." *Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge*. ACM, 2017.

Our Solution

1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-dependent
   Feature engineering

Our Solution

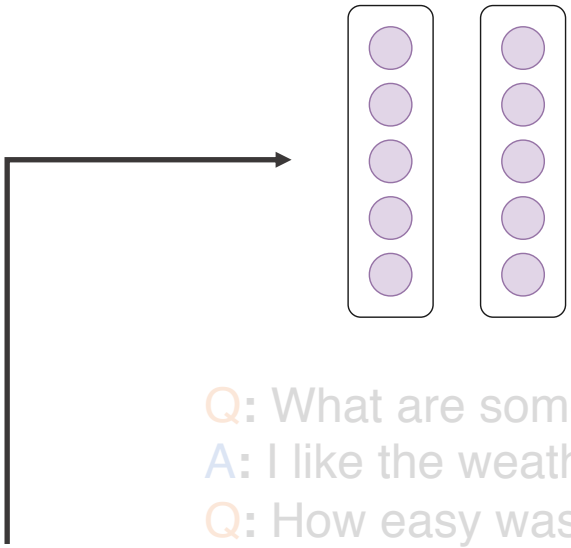
1. **Clinical:**
   Interview/questionnaire

2. **Automated**:
   Context-independent
   Feature engineering free

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Q: What'd you study at school?
A: um I took up business and administration.
Q: Cool are you still doing that?
A: Yeah I am. Here and there, I'm on a break right now but I plan on going back in the uh next semester.
Q: What's your dream job?
A: uh probably open up my own business.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Q: What'd you study at school?
A: um I took up business and administration.
Q: Cool are you still doing that?
A: Yeah I am. Here and there, I'm on a break right now but I plan on going back in the uh next semester.
Q: What's your dream job?
A: uh probably open up my own business.

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Q: What'd you study at school?
A: um I took up business and administration.
Q: Cool are you still doing that?
A: Yeah I am. Here and there, I'm on a break right now but I plan on going back in the uh next semester.
Q: What's your dream job?
A: uh probably open up my own business.



Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.
Q: What'd you study at school?
A: um I took up business and administration.
Q: Cool are you still doing that?
A: Yeah I am. Here and there, I'm on a break right now but I plan on going back in the uh next semester.
Q: What's your dream job?
A: uh probably open up my own business.

Method

…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

…

Method

1/0 depression

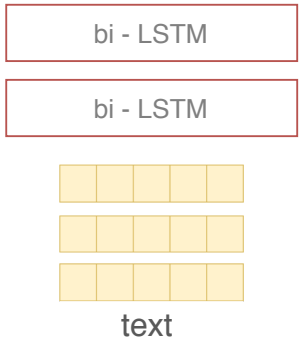
…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

…

Method

1/0 depression



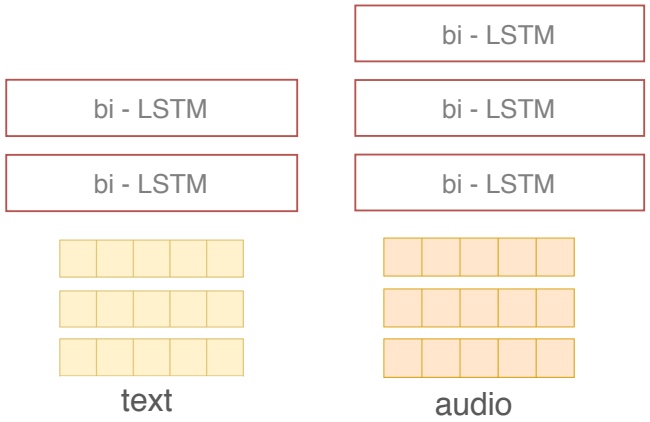
…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

…

Method

1/0 depression

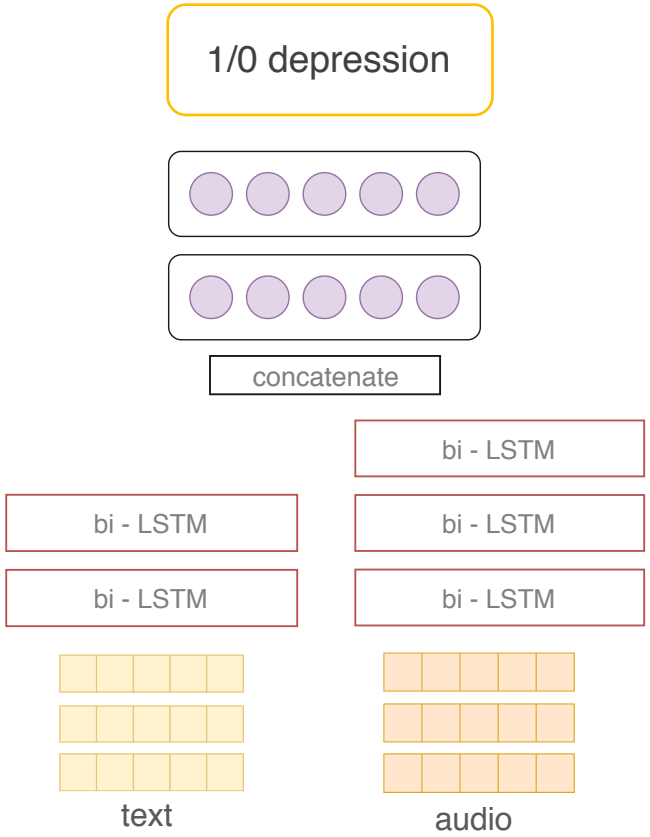


…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

…

Method



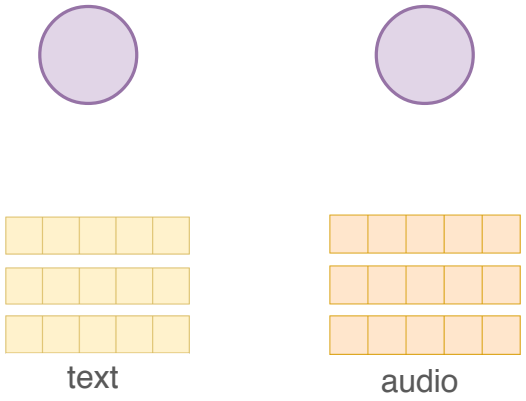
…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

…

Method (context-free)

1/0 depression



…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

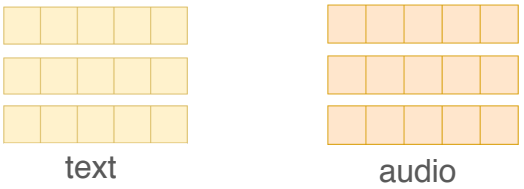
…

Method (context-free)

1/0 depression

P(depressed) P(depressed)

What do you like about LA? 0.2 0.4
How have you been feeling lately? 0.6 0.4
What's your sleep been like? 0.2 0.2



…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

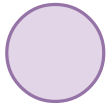
…

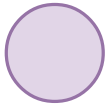
Method (weighted)

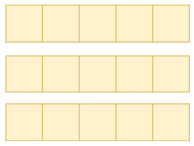
1/0 depression

What do you like about LA?
How have you been feeling lately?
What's your sleep been like?

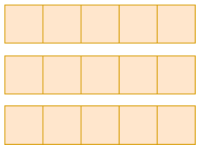
Q			Q		
0.4		0.2	0.4		0.4
0.5	x	0.6	0.7	x	0.4
0.3		0.2	0.2		0.2







text



audio

…

Q: What are some things you really like about L.A.?
A: I like the weather, I like the opportunities.
Q: How easy was it for you to get used to living in L.A.?
A: It took a minute, somewhat easy.
Q: What are some things you don't really like about L.A.?
A: Congestion.

…

Data

- **Corpus:** Distress Analysis Interview (DAIC).

Data

- **Corpus:** Distress Analysis Interview (DAIC).
- **Recordings:** 142 (train = 107, test = 35).

Data

- **Corpus:** Distress Analysis Interview (DAIC)
- **Recordings:** 142 (train = 107, test = 35).
- **Structure:** *Wizard-of-Oz* dialogue with 170 unique questions and 8,050 responses.

Data

- **Corpus:** Distress Analysis Interview (DAIC).
- **Recordings:** 142 (train = 107, test = 35).
- **Structure:** *Wizard-of-Oz* dialogue with 170 unique questions and 8,050 responses.
- **Outcome:** binary (28 depressed).

Features

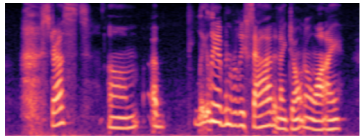
- **Audio:** (x 279)

Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

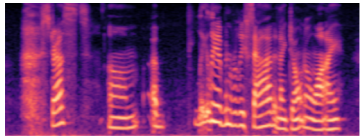
Features

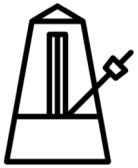
- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.



Features

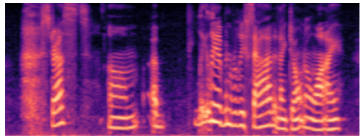
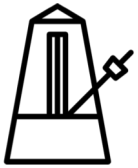
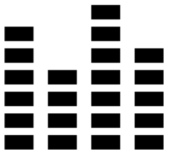
- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.





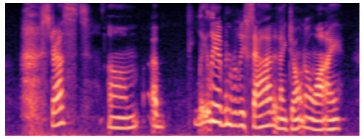
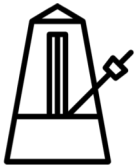
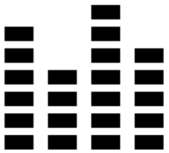
Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

Features

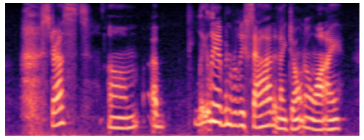
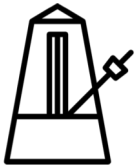
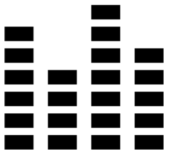
- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

Higher order statistics of response segment.
{mean, max, min, med, std, skew, kurt}

Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

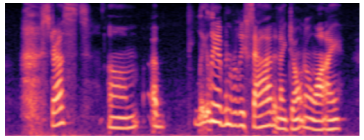
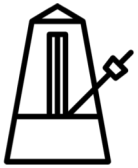
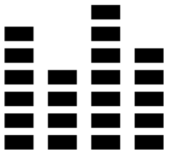




  Higher order statistics of response segment.
  {mean, max, min, med, std, skew, kurt}

- **Text:** (x 100)

Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

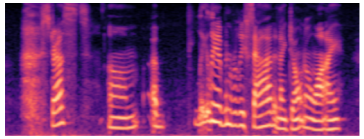
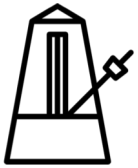
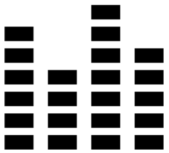




  Higher order statistics of response segment.
  {mean, max, min, med, std, skew, kurt}

- **Text:** (x 100)
  Doc2Vec embeddings of question-response.

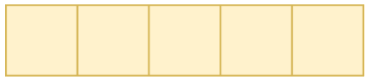
Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

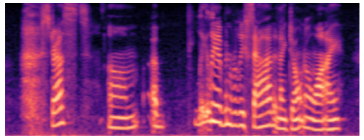
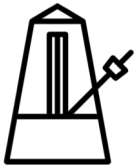
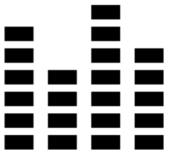
  Higher order statistics of response segment.
  {mean, max, min, med, std, skew, kurt}

- **Text:** (x 100)
  Doc2Vec embeddings of question-response.



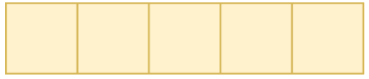
Features

- **Audio:** (x 279)
  Spectral energy, prosody, and voice quality.

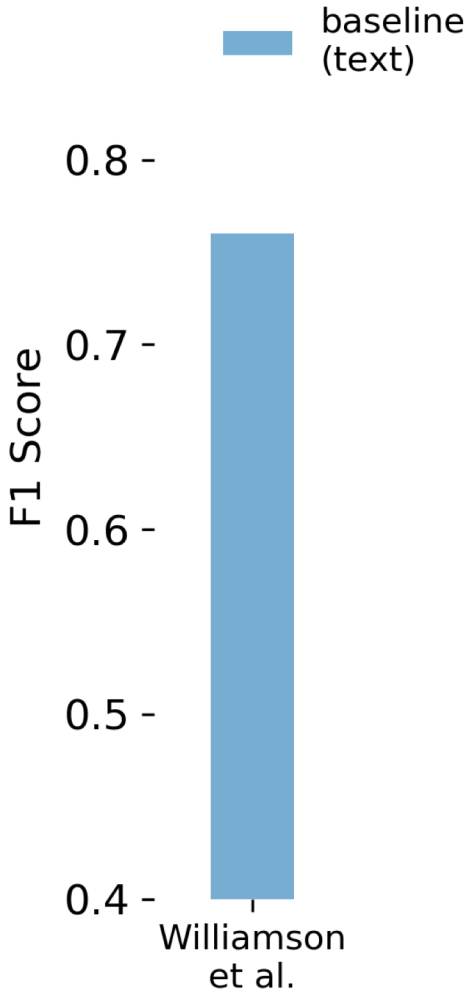
  Higher order statistics of response segment.
  {mean, max, min, med, std, skew, kurt}

- **Text:** (x 100)
  Doc2Vec embeddings of question-response.

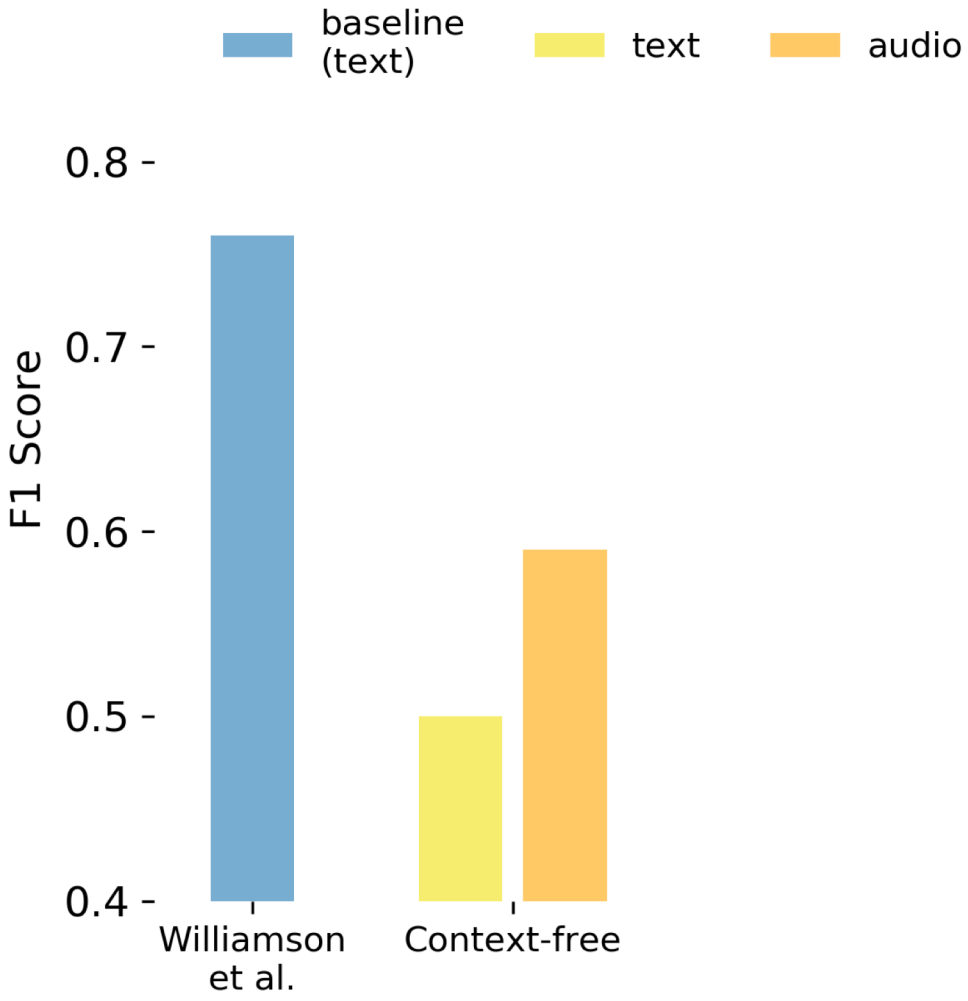


  {dim = 100, min-words = 3,
  context-win = 3, epochs = 50}

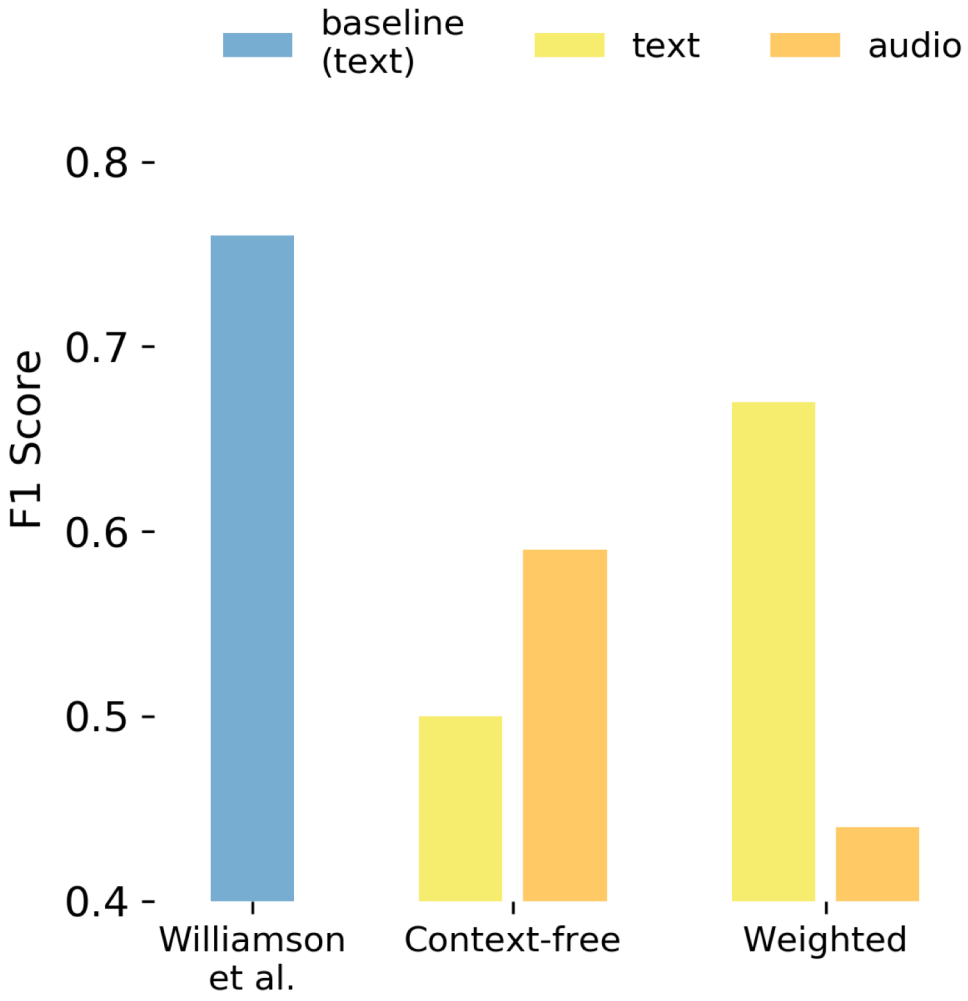
Results



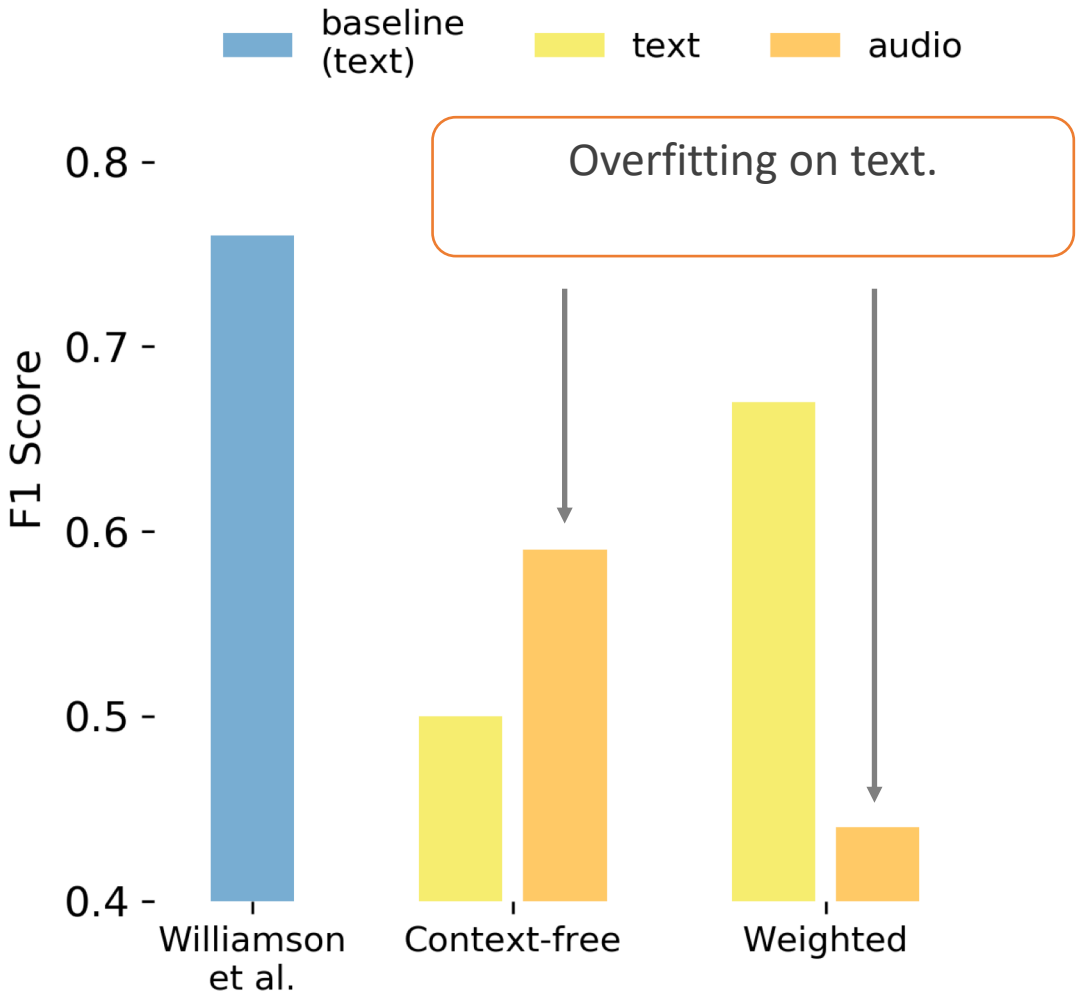
Results


baseline (text)
text
audio
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free

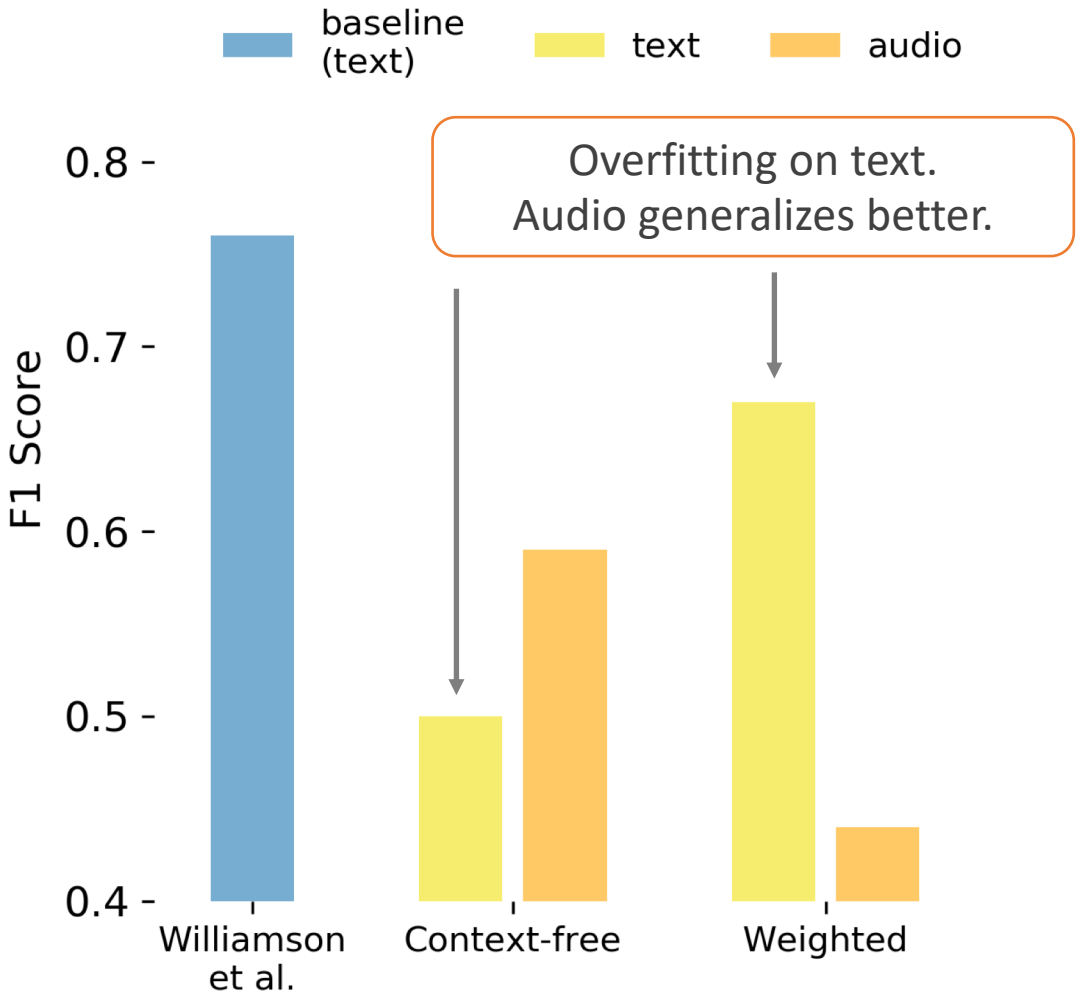
Results


baseline (text)
text
audio
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted

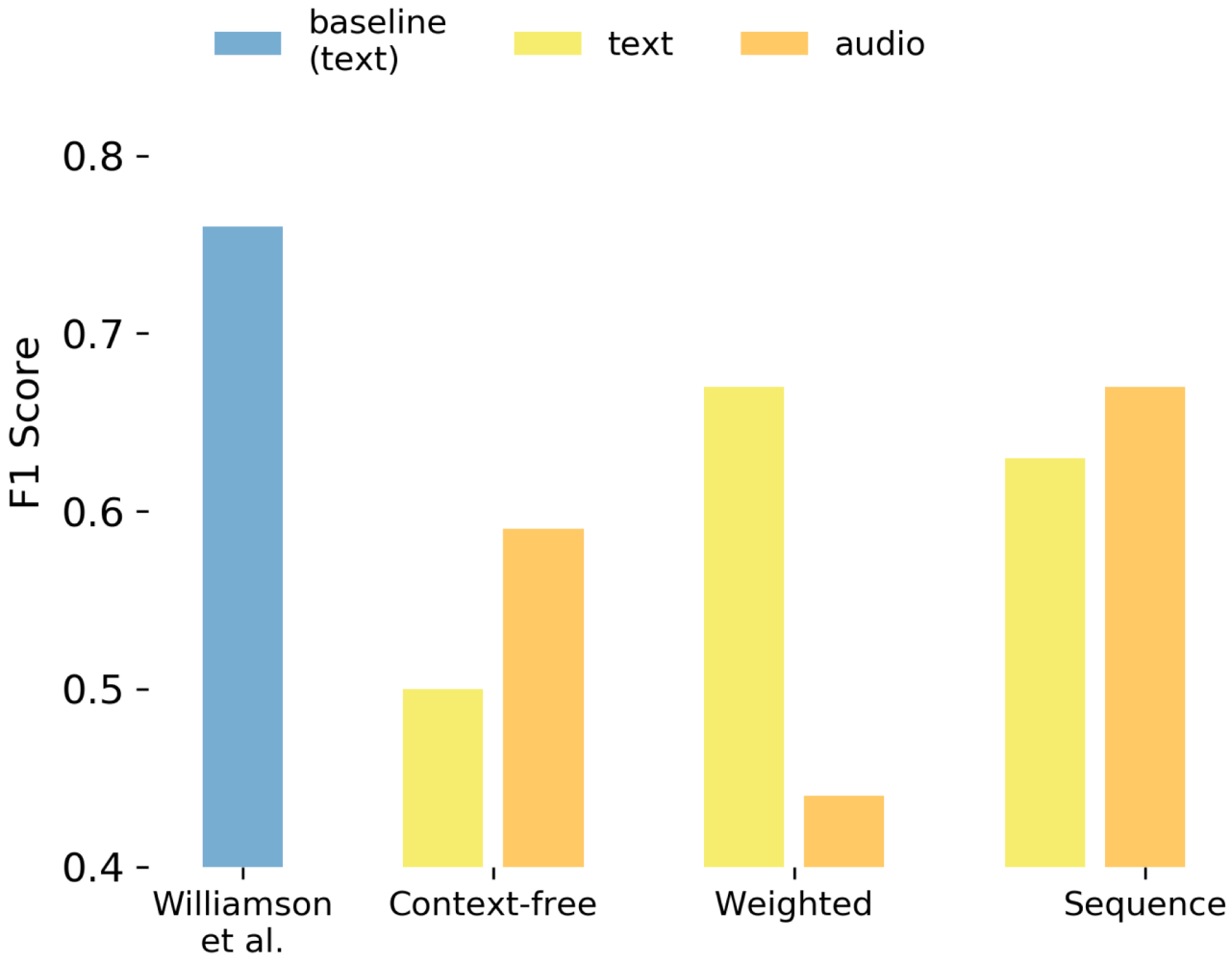
Results


baseline (text)
text
audio
Overfitting on text.
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted

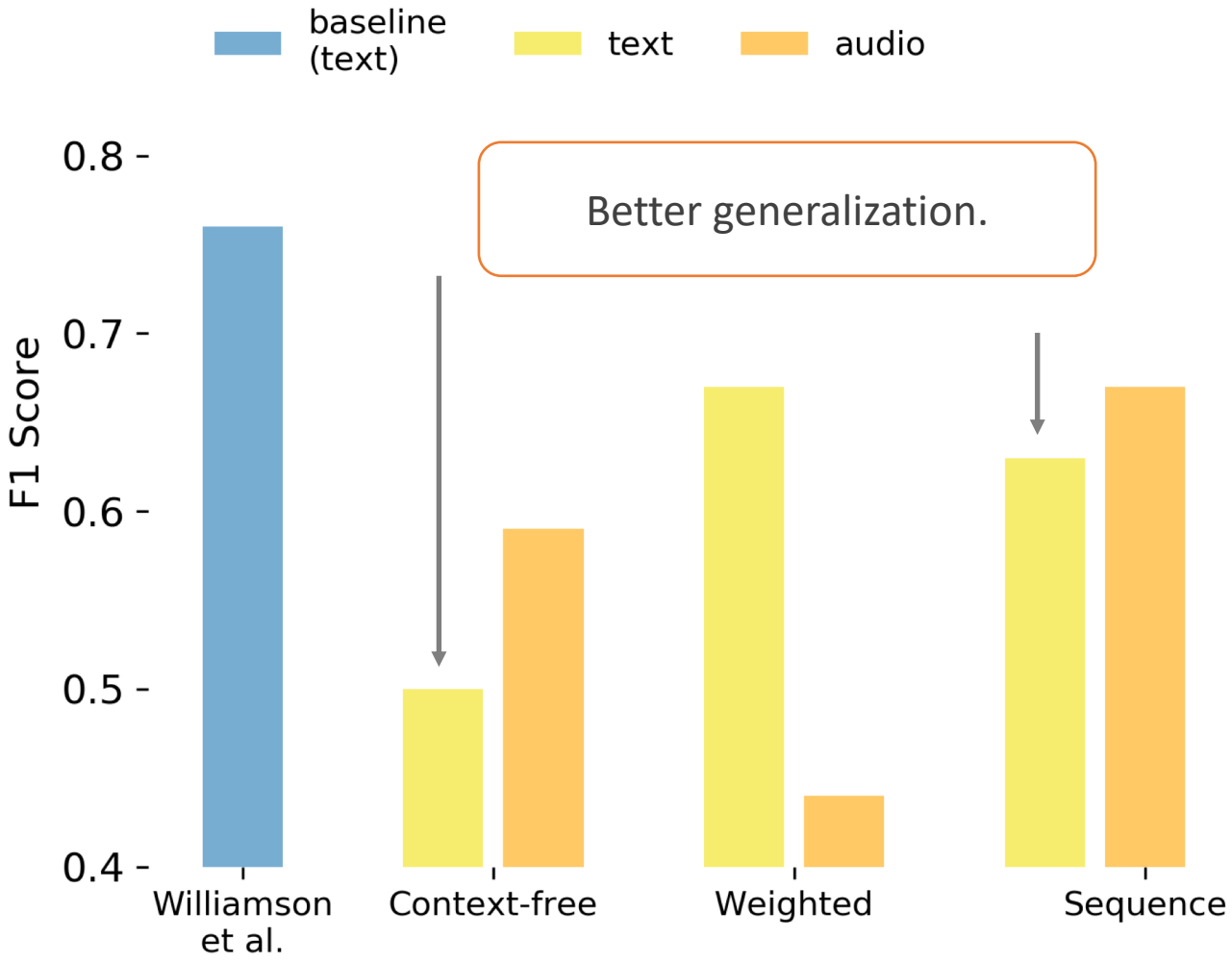
Results


baseline (text)
text
audio
Overfitting on text.
Audio generalizes better.
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted

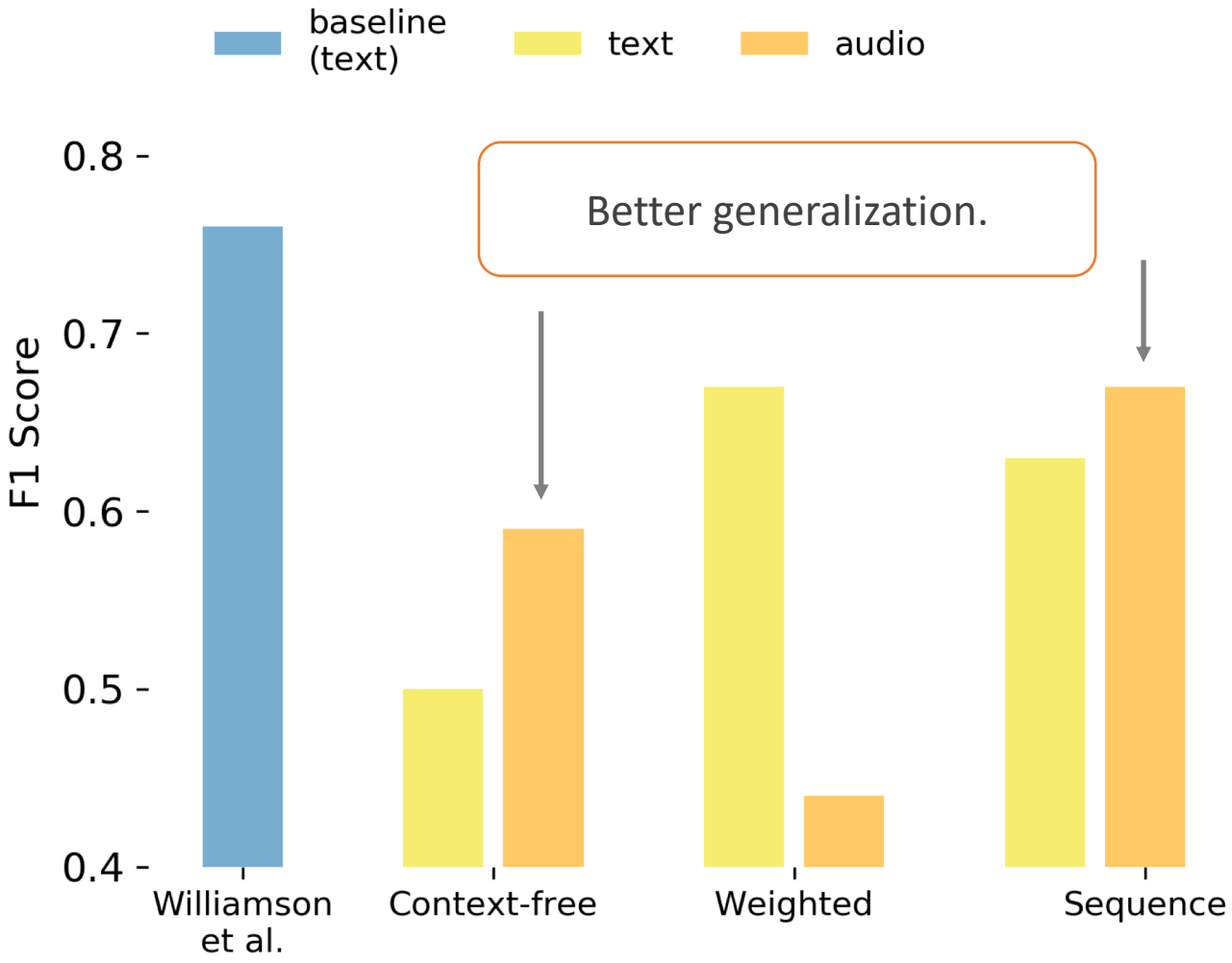
Results


baseline (text)
text
audio
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

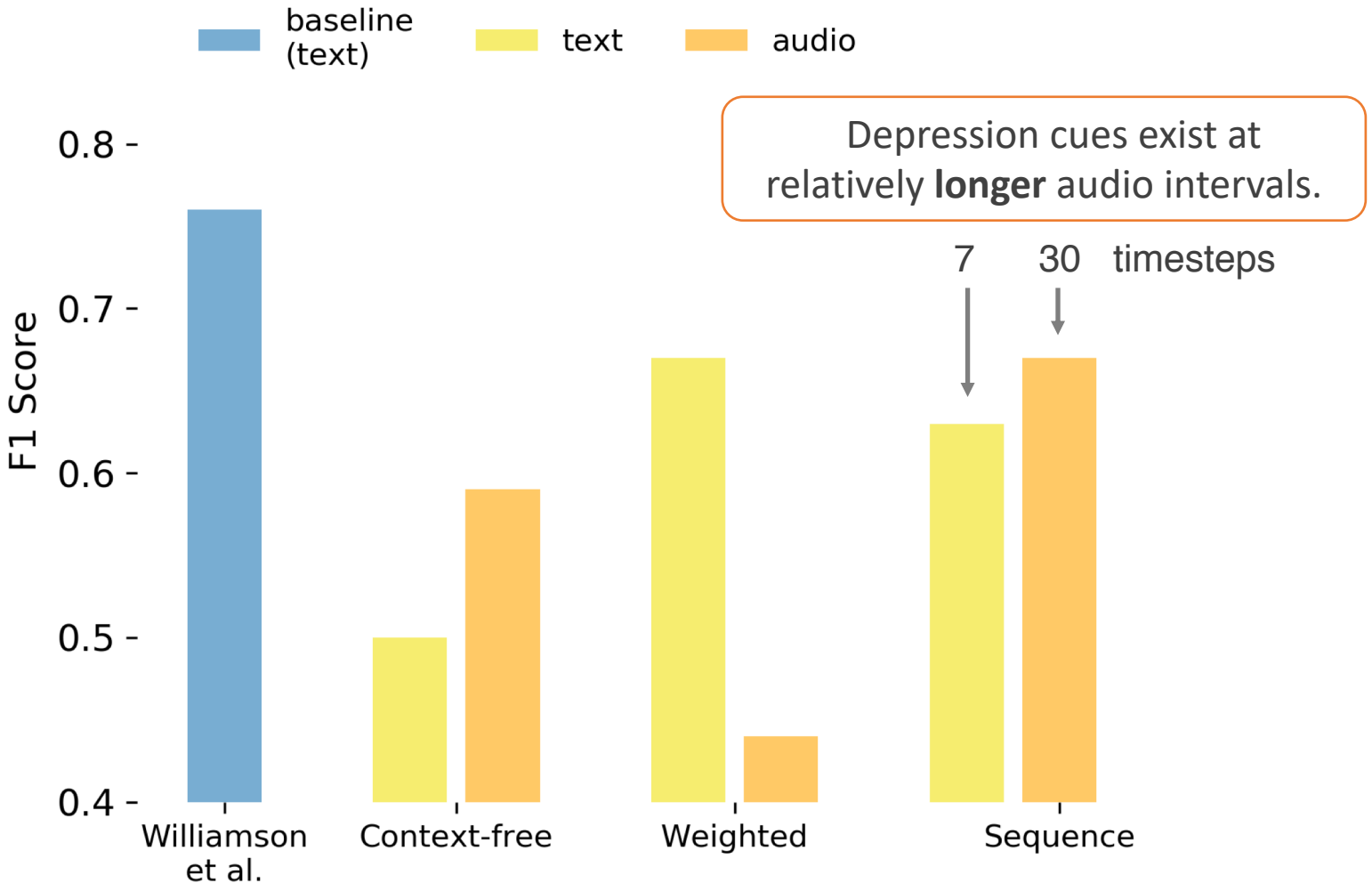
Results


baseline (text)
text
audio
Better generalization.
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

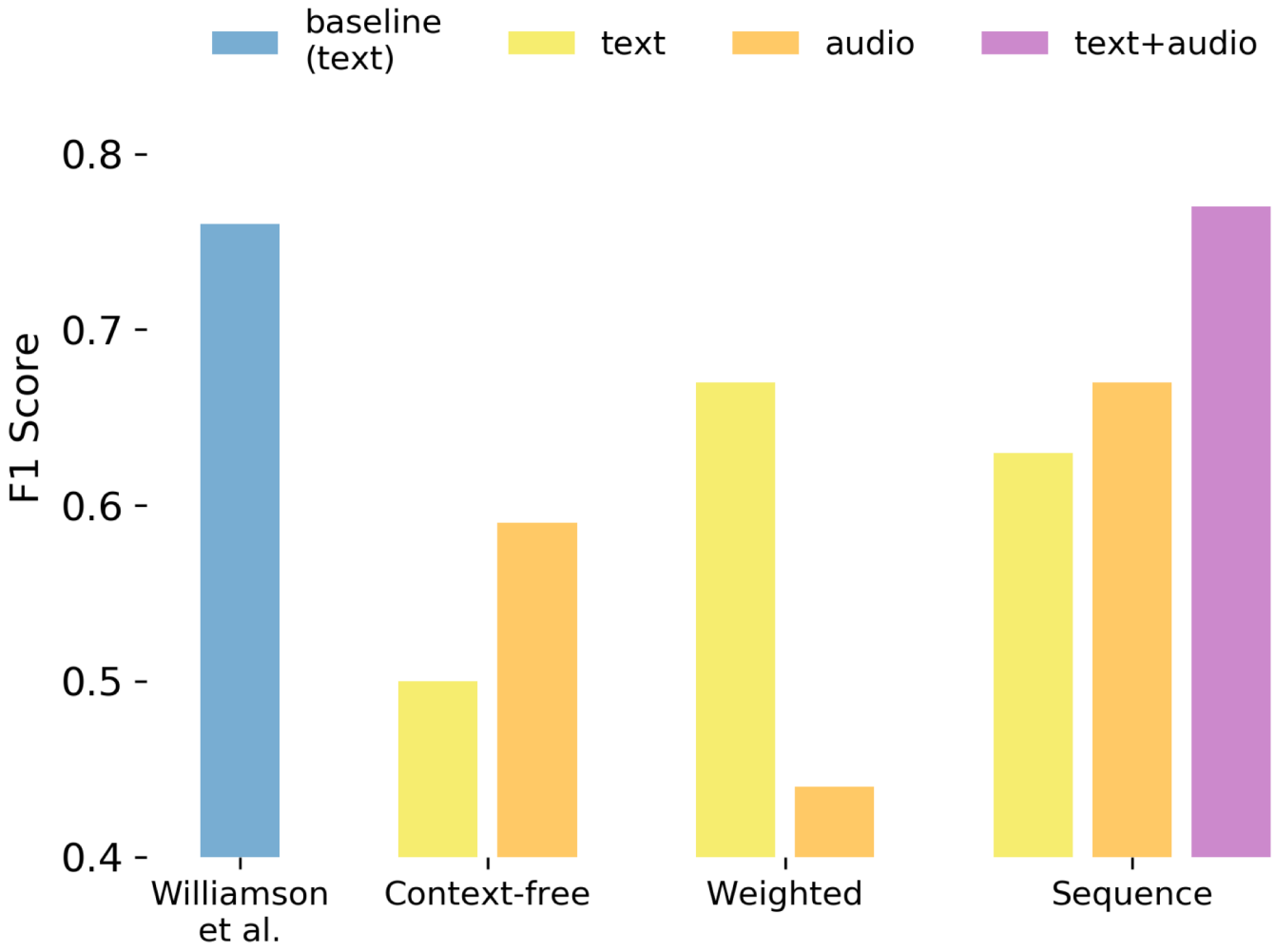
Results


baseline (text)
text
audio
Better generalization.
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

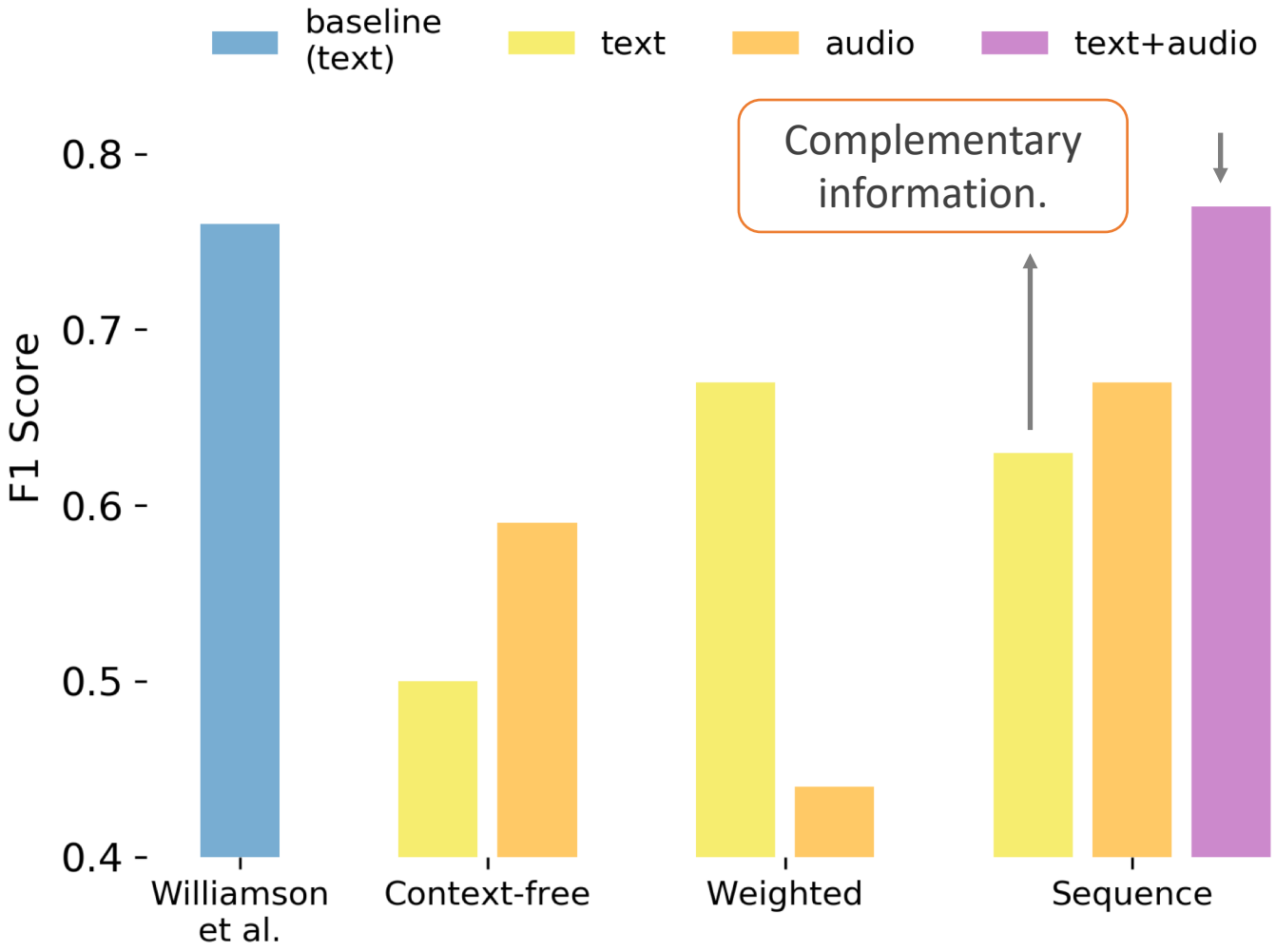
Results


baseline (text)
text
audio
Depression cues exist at relatively **longer** audio intervals.
7
30
timesteps
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

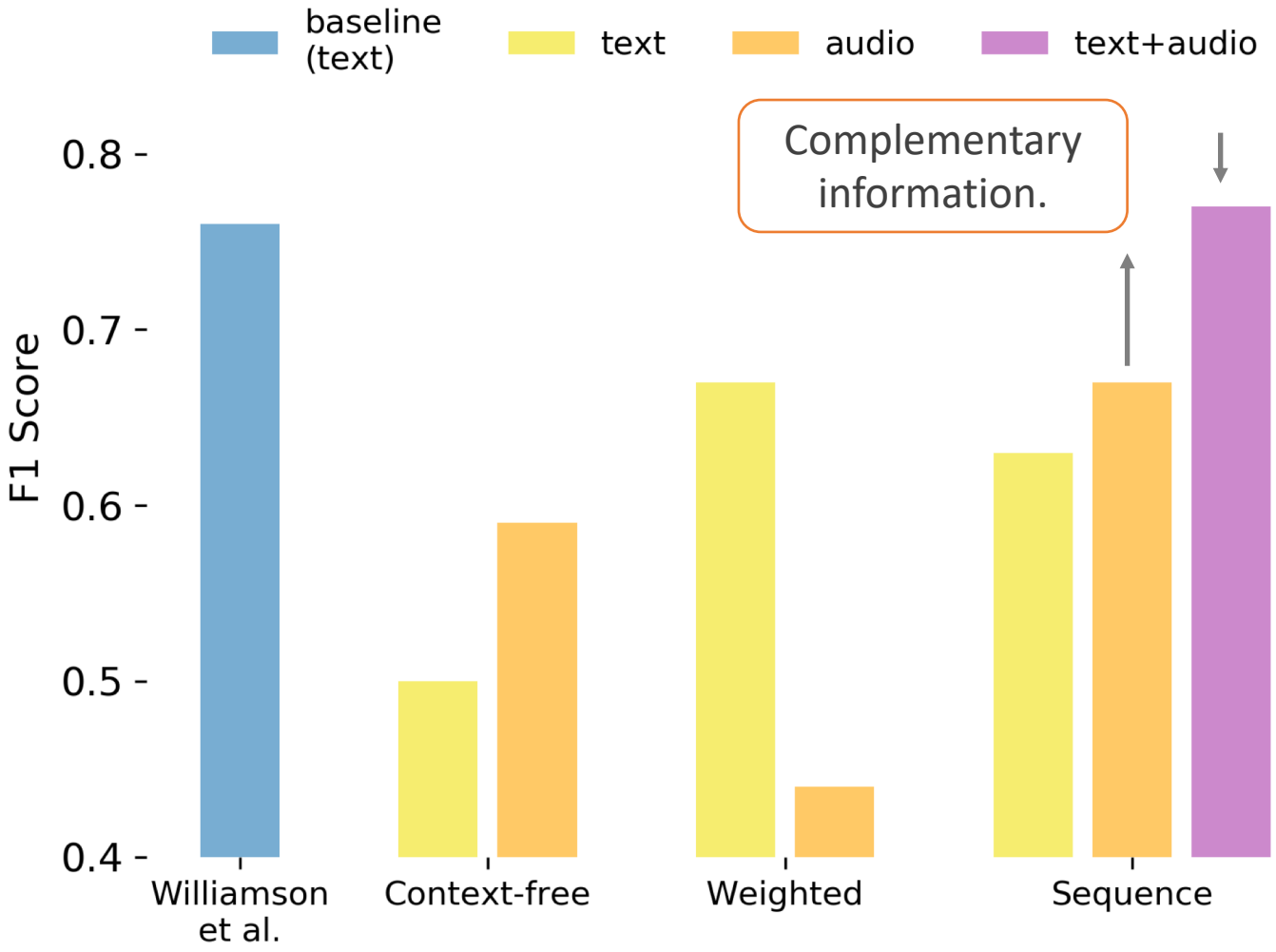
Results


baseline (text)
text
audio
text+audio
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

Results


baseline (text)
text
audio
text+audio
Complementary information.
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

Results


baseline (text)
text
audio
text+audio
Complementary information.
F1 Score
0.8
0.7
0.6
0.5
0.4
Williamson et al.
Context-free
Weighted
Sequence

Conclusion

1. **Sequence Modeling:**
   Improved generalization.
   Better than baseline(s).

2. **Modality Inputs to Model:**
   audio = 30 sequences
   text = 7 sequences
   Depression cues exist at longer speech intervals.

3. **Weighted Model:**
   Overfitting on text.

Future Work

1. Apply technique to larger number of subjects with different conditions (dementia).

2. Infer most predictive segments.

3. Infer patterns being captured (speaking rate? keywords?)

Reference

T. Alhanai, MM. Ghassemi, and J. Glass, "Detecting Depression with Audio/Text Sequence Modeling of Interviews," Proc. Interspeech, Hyderabad, India, September 2018.

Publication:
https://groups.csail.mit.edu/sls/publications/2018/Alhanai_Interspeech-2018.pdf

Github:
https://github.com/talhanai/redbud-tree-depression