WINTERLIGHT

Multimodal deep-learning system for anxiety and depression detection

Brian Diep, Marija Stanojevic, Jekaterina Novikova

Overview

- Motivation
- Dataset
- Model
- Results
- Conclusion

Motivation

- Mental health care is an integral part of providing holistic care for patients
- Mental health screening remains a barrier to many who wish to access mental health care
- Improvements in the diagnosis or monitoring process can lead to huge benefits to those who suffer from depression and anxiety

WINTERLIGHT

WHY SPEECH?

WHY SPEECH?

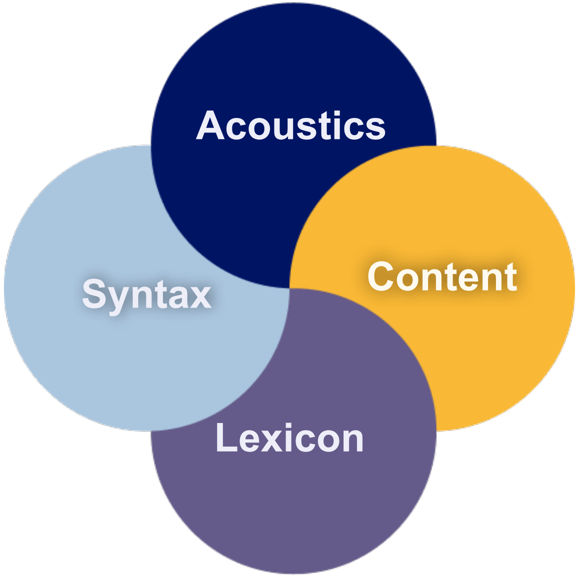
Speech is a rich source of information about
disease severity & progression.

- Alzheimer's Disease
- Depression
- Schizo-phrenia
- Anxiety
- Fronto-temporal dementia

Disease states produce measurable changes in rate of speech, number of pauses, amount of detail provided, and types of words used.

SPEECH ASSESSMENTS

Automated speech analysis simultaneously assesses different **domains of speech and language**.



Benefits of using speech:

- Ecologically valid
- Low patient burden
- Functionally relevant
- Can be assessed remotely and at high frequency
- Objective

DATA COLLECTION

Dataset

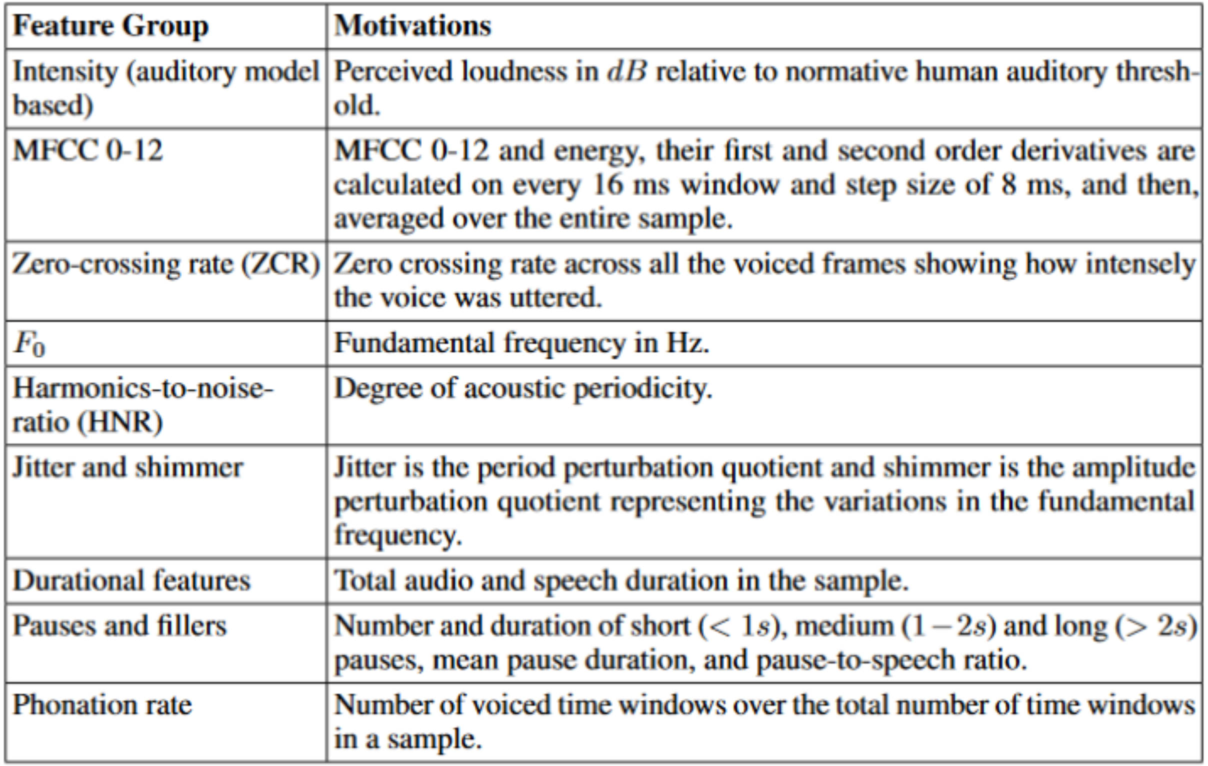
- Data sourced from the DEPAC corpus, collected over Amazon mTurk
- Individuals were prompted to record themselves completing self-administered speech tasks
- These were then automatically transcribed to form an associated transcript for each audio file
- Individuals also complete a questionnaire including demographic details and questions following the GAD-7 and PHQ-8 scale

Mashrura Tasnim, Malikeh Ehghaghi, Brian Diep, and Jekaterina Novikova. 2022. DEPAC: a Corpus for Depression and Anxiety Detection from Speech. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 1–16, Seattle, USA. Association for Computational Linguistics.

Dataset Tasks

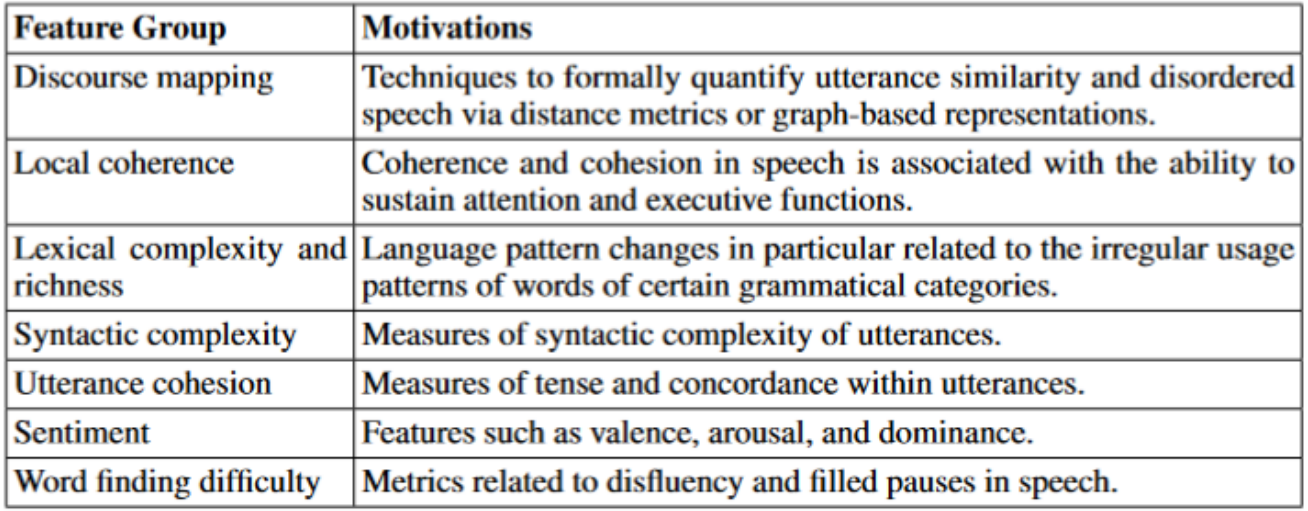
- Journaling
  - Subjects are asked to describe their day in as much detail as they would like
- Prompted narrative
  - Subjects are asked to describe hobbies or travel experiences with as much details as they would like
- Positive fluency
  - Subjects are asked to list as many positive events that they expect to experience in the near future

Dataset Features



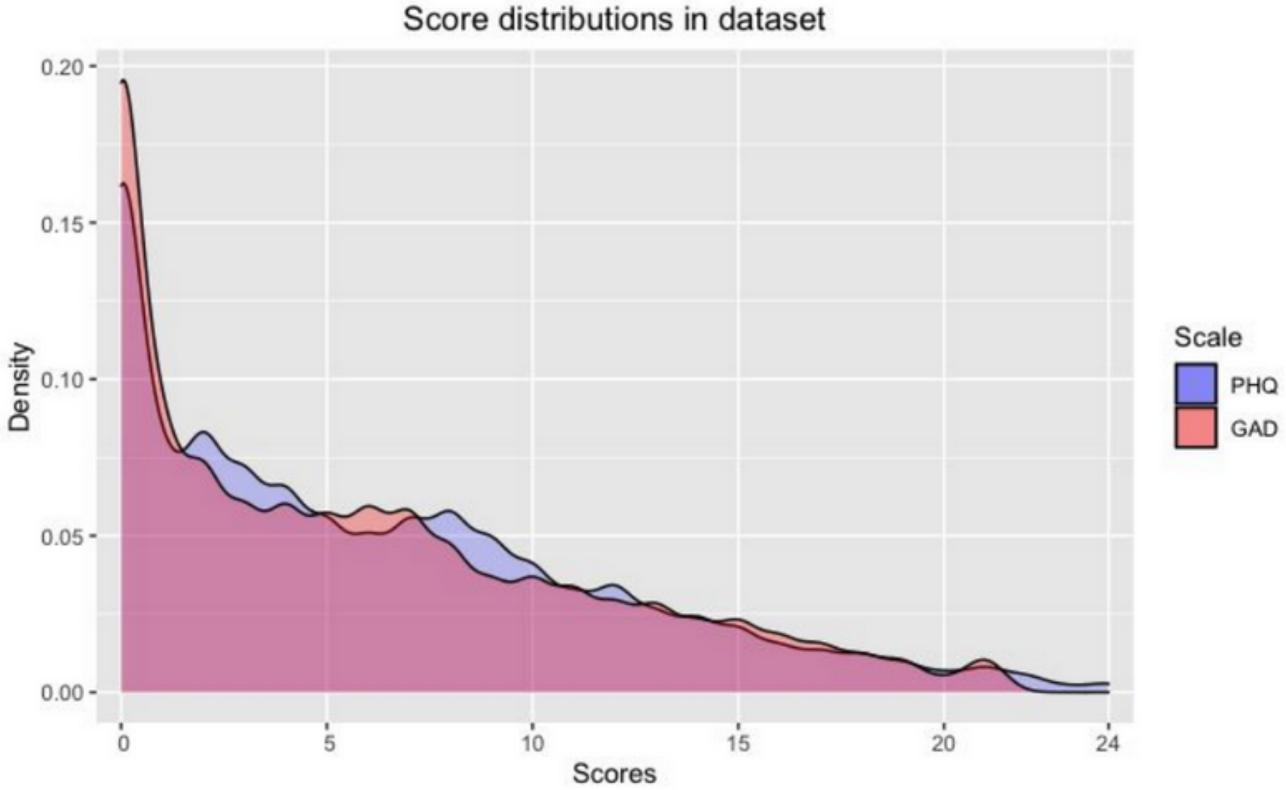
Feature Group	Motivations
Intensity (auditory model based)	Perceived loudness in dB relative to normative human auditory threshold.
MFCC 0-12	MFCC 0-12 and energy, their first and second order derivatives are calculated on every 16 ms window and step size of 8 ms, and then, averaged over the entire sample.
Zero-crossing rate (ZCR)	Zero crossing rate across all the voiced frames showing how intensely the voice was uttered.
F_0	Fundamental frequency in Hz.
Harmonics-to-noise-ratio (HNR)	Degree of acoustic periodicity.
Jitter and shimmer	Jitter is the period perturbation quotient and shimmer is the amplitude perturbation quotient representing the variations in the fundamental frequency.
Durational features	Total audio and speech duration in the sample.
Pauses and fillers	Number and duration of short ($< 1s$), medium ($1 - 2s$) and long ($> 2s$) pauses, mean pause duration, and pause-to-speech ratio.
Phonation rate	Number of voiced time windows over the total number of time windows in a sample.

Dataset Features



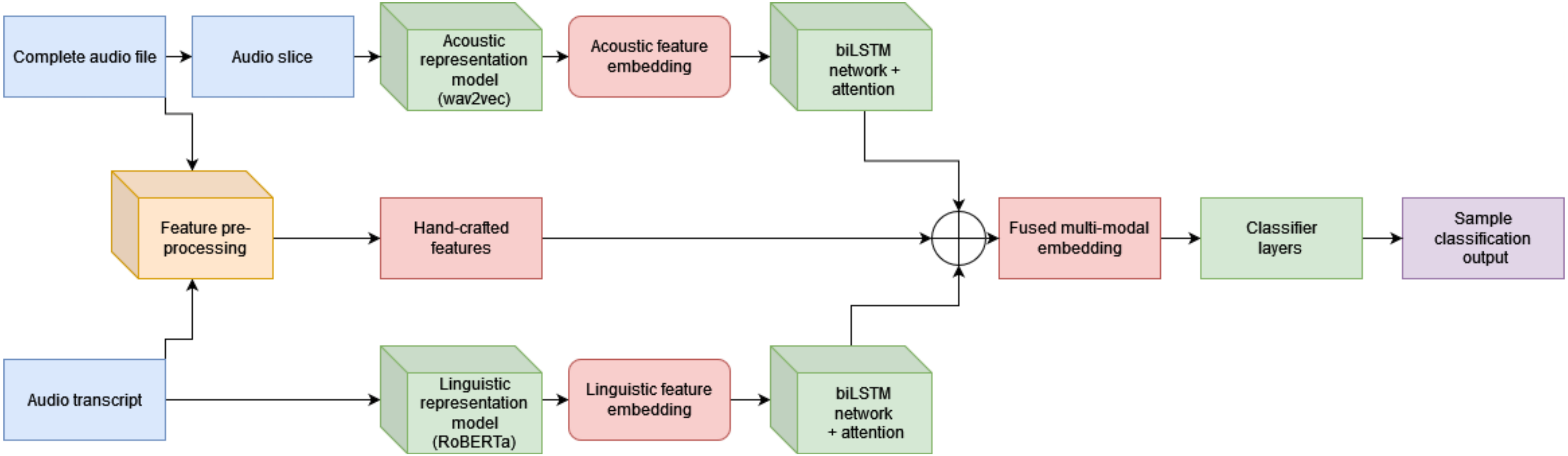
Feature Group	Motivations
Discourse mapping	Techniques to formally quantify utterance similarity and disordered speech via distance metrics or graph-based representations.
Local coherence	Coherence and cohesion in speech is associated with the ability to sustain attention and executive functions.
Lexical complexity and richness	Language pattern changes in particular related to the irregular usage patterns of words of certain grammatical categories.
Syntactic complexity	Measures of syntactic complexity of utterances.
Utterance cohesion	Measures of tense and concordance within utterances.
Sentiment	Features such as valence, arousal, and dominance.
Word finding difficulty	Metrics related to disfluency and filled pauses in speech.

Dataset


Score distributions in dataset
Density
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
24
Scores
Scale
PHQ
GAD

MODEL ARCHITECTURE

Model Architecture

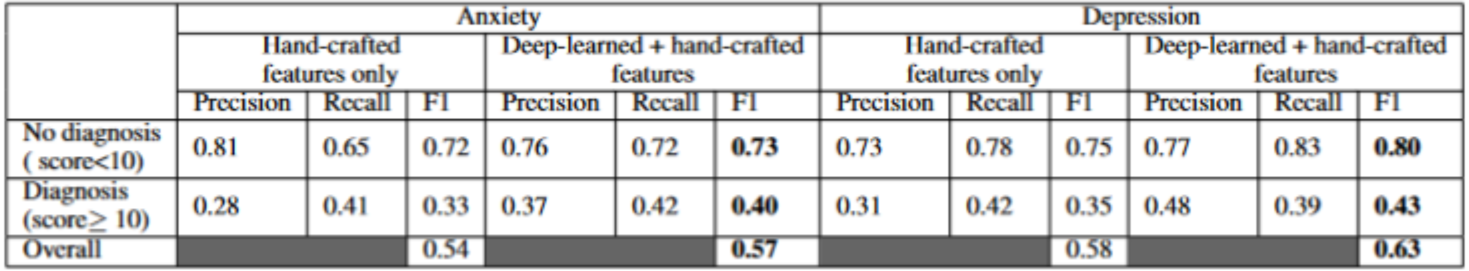

Complete audio file
Audio slice
Acoustic representation model (wav2vec)
Acoustic feature embedding
biLSTM network + attention
Feature pre-processing
Hand-crafted features
Fused multi-modal embedding
Classifier layers
Sample classification output
Audio transcript
Linguistic representation model (RoBERTa)
Linguistic feature embedding
biLSTM network + attention

extends work presented in AudiBERT - Toto, Ermal, M. L. Tlachac, and Elke A. Rundensteiner. "Audibert: A deep transfer learning multimodal classification framework for depression screening." *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021.

WINTERLIGHT

RESULTS

Results



	Anxiety						Depression					
	Hand-crafted features only			Deep-learned + hand-crafted features			Hand-crafted features only			Deep-learned + hand-crafted features		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
No diagnosis (score<10)	0.81	0.65	0.72	0.76	0.72	**0.73**	0.73	0.78	0.75	0.77	0.83	**0.80**
Diagnosis (score≥ 10)	0.28	0.41	0.33	0.37	0.42	**0.40**	0.31	0.42	0.35	0.48	0.39	**0.43**
Overall			0.54			**0.57**			0.58			**0.63**

- The addition of hand-crafted features improves our performance for anxiety and depression classification over the baseline
- This is reflective of existing work that shows that the addition of language models like BERT can improve depression classification

Discussion

- There is an effect caused by data imbalance between the diagnosis and no diagnosis classes
    - Most data falls under the diagnosis cutoff (only 12.8% and 25.3% of the anxiety and depression samples respectively have scores above the cutoff)
- Depression classification overall has higher performance than anxiety

WINTERLIGHT

CONCLUSION

Conclusion

- Speech is an effective modality for the diagnosis of depression and anxiety
- There is value in combining deep-learned and hand-crafted features for depression and anxiety detection
- Machine learning and crowdsourcing pose a new and exciting opportunity to potentially improve mental health care, and make it more accessible to all

WINTERLIGHT

Johnson & Johnson INNOVATION | JLABS
JLABS @ Toronto

MaRS

hv.
hikma ventures

GREYSKY
venture partners

First Star Ventures™

NOVATIO

Ontario Bioscience BIO

UNIVERSITY OF TORONTO

ONTARIO
BRAIN
INSTITUTE