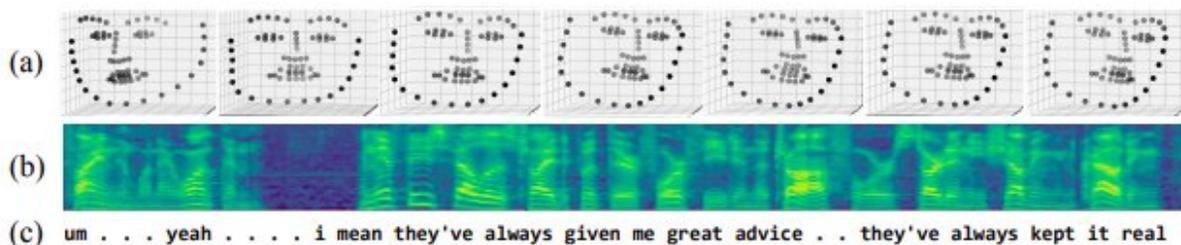


# Analyzing spoken language and 3-D facial expressions to measure depression severity

December 4 2018, by Ingrid Fadelli

---



Multi-modal data. For each clinical interview, the researchers use: (a) video of 3D facial scans, (b) audio recording, visualized as a log-mel spectrogram, and (c) text transcription of the patient's speech. The model predicts the severity of depressive symptoms using all three modalities. Credit: Haque et al.

Researchers at Stanford have recently explored the use of machine learning to measure the severity of depressive symptoms by analyzing people's spoken language and 3-D facial expressions. Their multi-model method, outlined in a paper pre-published on arXiv, achieved very promising results, with an 83.3 percent sensitivity and 82.6 percent specificity.

Currently, over 300 million people worldwide suffer from depression disorders to varying degrees. In extreme cases, depression can lead to suicide, with an average of approximately 800,000 people committing

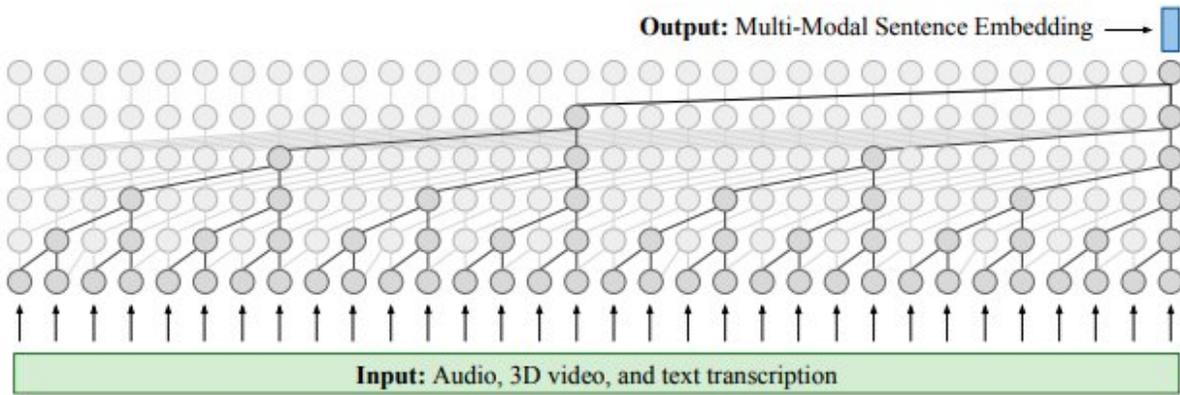
suicide every year.

Mental health disorders are currently diagnosed upon careful examination by a wide range of health care providers, including primary care physicians, clinical psychologists and psychiatrists. Nonetheless, detecting mental illnesses is often far more challenging than diagnosing physical illnesses.

Several factors, including [social stigma](#), treatment cost and availability, might prevent affected individuals from seeking help. Currently, researchers estimate that 60 percent of those affected by [mental illnesses](#) do not receive treatment.

Developing methods that can automatically detect [depressive symptoms](#) could improve the accuracy and availability of diagnostic tools, leading to faster and more efficient interventions. A team of researchers at Stanford have recently investigated the use of machine learning to measure the severity of depressive symptoms.

"In this work, we present a machine learning method for measuring the severity of depressive symptoms," the researchers wrote in their paper. "Our multi-modal method uses 3-D facial expressions and spoken language, commonly available from modern cell phones."



Learning a multi-modal sentence embedding. Overall, the model is a causal CNN. The input for the model is: audio, 3D facial scans, and text. The multi-modal sentence embedding is fed to a depression classifier and PHQ regression model (not shown above). Credit: Haque et al.

Depressed individuals often present a series of verbal and non-verbal symptoms, including monotone pitch, reduced articulation rate, lower speaking volumes, fewer gestures, and more downward gazes. One of the most common tests to assess the severity of depression symptoms is the patient health questionnaire (PHQ).

The method devised by the researchers analyzes audio tracks of patients' voice, 3-D video of their [facial expressions](#), and text transcriptions of their clinical interviews. Based on this data, the model produces either a PHQ score or classification label indicating major depressive disorder.

In an initial evaluation, the model achieved an average error of 3.67 points (15.3 percent relative), on the PHQ scale, detecting major depressive disorder with 83.3 percent sensitivity and 82.6 percent specificity. The researchers chose to collect the data used in their study via human-to-computer interviews, rather than human-to-human ones.

"Compared to a human interviewer, research has shown that patients report lower fear of disclosure and display more [emotional intensity](#) when conversing with an avatar," the researchers wrote. "Additionally, people experience psychological benefits from disclosing emotional experiences to chatbots."

In the future, this new machine learning method could be deployed in smartphones worldwide, aiding the mission of making mental health care cheaper and more accessible. According to the researchers, their model is designed to augment and complement existing clinical methods, rather than issuing formal diagnoses.

"We presented a multi-modal machine learning [method](#) which combines techniques from [speech recognition](#), computer vision, and natural language processing," the researchers wrote. "We hope this work will inspire others to build AI-based tools for understanding [mental health disorders](#) beyond depression."

**More information:** Measuring depression symptom severity from spoken language and 3D facial expressions. arXiv:1811.08592 [cs.CV]. [arxiv.org/abs/1811.08592](https://arxiv.org/abs/1811.08592)

© 2018 Science X Network

Citation: Analyzing spoken language and 3-D facial expressions to measure depression severity (2018, December 4) retrieved 26 April 2024 from <https://techxplore.com/news/2018-11-spoken-language-d-facial-depression.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.