Detecting Depression using Speech Processing

Pranav Bijith, Farhan Ansari

Background

- 20.78% of individuals in the U.S. are depressed from the period 2018-2019.
- Tools that help doctors identify patients that could potentially have depression would be useful in helping people get the treatment they need.
- Depression has effects on a patient's voice, often lowering their pitch and increasing the length of pauses.
- The DAIC-WoZ dataset contains 189 voice recordings of depressed and non-depressed patients talking to a human-controlled virtual interviewer.

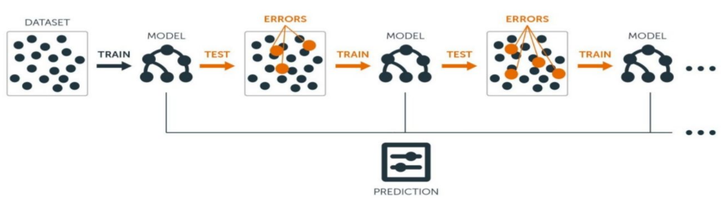
Research Question

How can we detect depression using speech recognition?

Hypothesis

- We predict that our model will be mediocre at average around a 60-65% due to a lack of computing power and state of the art models being around 70-75%
- We predict that gradient boosting be more effective compared to a CNN

Methods



- We used a gradient boosting algorithm coupled with feature extraction from the library librosa. There were 68 features extracted, including pitch, mfcc, spectral centroid, etc.
- For gradient boosting, we used SMOTE, cross validation, and K-Best optimize the model
- For our second model, we used a CNN with 4 layers of convolution paired with dropout and max pooling. It also had 2 Relu layers and a softmax layer at the end.

Results

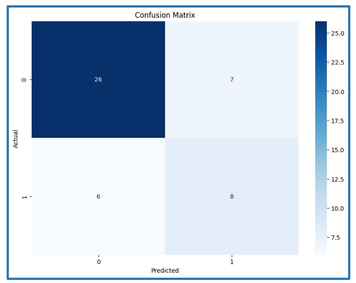


Fig. 1 (59 feat.)

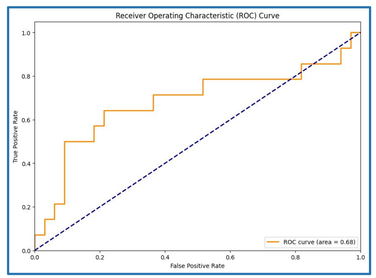


Fig. 2 (59 feat.)

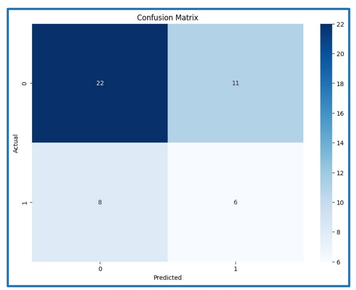


Fig. 3 (20 feat.)

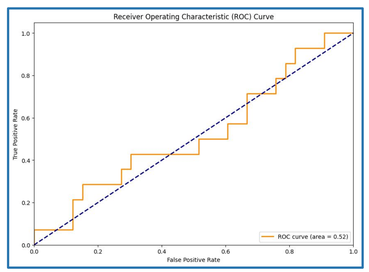


Fig. 4 (20 feat.)

Table 1: Results for precision, recall, F1-Score and Support for Gradient Boosting model

Class	Precision	Recall	F1-Score	Support
Not Depressed	0.81	0.79	0.80	33
Depressed	0.53	0.57	0.55	14

Table 2: List of Parameters used in the Gradient Boosted Folds

n_estimators	learning_rate	max_depth	min_samples_split
340.0	0.01	13	2
345.0	0.05	15	5
350.0	0.07	17	7
355.0	0.08	20	9
360.0	0.085	25	25
370.0	0.09	30	30
330.0	0.095	45	45
	0.12	60	55
		100	75

Table 3: Accuracy, Macro Average, and Weighted Average for Gradient Boosting model

Accuracy	0.72
Macro Avg	0.68
Weighted Avg	0.73

Discussion

- As seen in Figures 1 & 2, using 59 features was more suitable than using 20 features as demonstrated in Figures 3 & 4.
- The macro F1 score for this model is 0.675 and the weighted accuracy was 73% (Table 3), which was around the low end of current state-of-the-art models, which use more advanced techniques.

Method	Paper	Dataset	Performance
LSTM	Alhanai et al. 2018 [37]	DAIC-WOZ	MAE/RMSE 4.97/6.27
	Du et al. 2018 [88]	BD	UAR/UAP/Accuracy 0.651/0.678/65.0%
	Salekin et al. 2018 [89]	DAIC-WOZ	F1/Accuracy 0.901/90%
	Othmani et al. 2021 [77]	DAIC-WOZ	F1 (D/N) 0.49/0.82 accuracy 73.35%
	Zhang et al. 2021 [38]	DAIC-WOZ	MAE/RMSE 5.48/6.31
CNN	Yang et al. 2017 [84]	DAIC-WOZ	MAE/RMSE 5.163/5.974
	Haque et al. 2018 [78]	DAIC-WOZ	F1/Precision/Recall 0.769/71.4%/83.3%
	He et al. 2018 [79]	AVEC2013/14	MAE/RMSE 8.78/10.90
	Huang et al. 2020 [44]	DAIC-WOZ	F1/Accuracy 0.700/82.9%
	Muzammel et al. 2020 [83]	DAIC-WOZ	Accuracy/Precision/ Recall/F1 86.06%/81%/73%/77%
	Vázquez-Romero et al. 2020 [85]	DAIC-WOZ	F1/Accuracy/Precision/ Recall 0.65/74%/55%/79%

- Some errors came from patients with minor depression.
- Our hypothesis that the gradient boosting model would outperform the CNN was correct. The CNN model only predicted "Not depressed", making it ineffective as a model despite its high accuracy.
- The gradient boosting algorithm was also not entirely fine tuned as seen with the hyperparameter distribution in Table 2.
- 59 features were used due to it yielding the greatest result with the parameters used in the Gradient Boosted folds

Applications

- Machine learning models that can use speech to detect depression could be valuable for physicians to get patients treatment. Having a software that could reliably detect depression would be able to diagnose many people who would have otherwise never been treated.

Future Exploration

- How can models be revised to maintain patients' privacy?
- Would integrating NLP to analyze the content of a patient's speech make the model more effective?

References

- Santilli, Mara. "Depression Statistics in 2023." Forbes Health, 5 July 2023, www.forbes.com/health/mind/depression-statistics/.
- Wu, Pingping et al. "Automatic depression recognition by intelligent speech signal processing: A systematic survey" 21 June 2022, https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/cit2.12113
- Kumar, Ajitesh. "Gradient Boosting Algorithm: Concepts, Example." Data Analytics, 3 Dec. 2022, vitalflux.com/gradient-boosting-algorithm-concepts-example/.