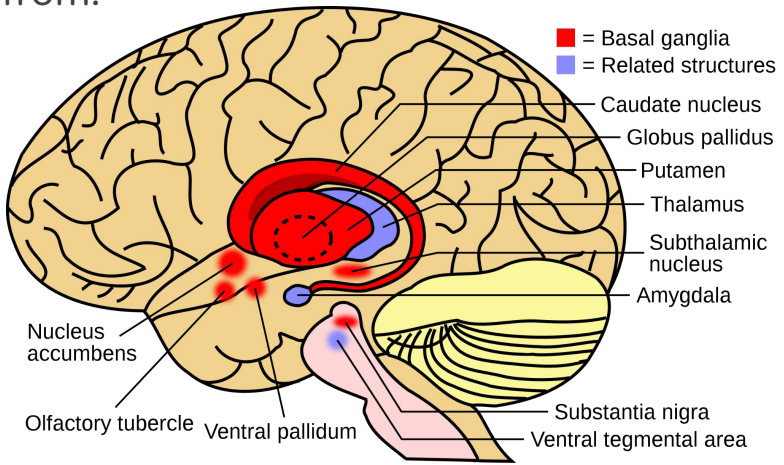
A Machine Learning Approach To Detecting Neurological Disorders

Detection of Degenerative Conditions via Audio Analysis

Background Info

- Our project aims to detect whether or not a person has a neurological disorder just by analyzing their voice.
- We looked at people with Spasmodic Dysphonia and ALS
- People with neurological disorders can suffer from:
  - slurred or slow speech
  - shaky voice
  - breathy speech
  - muscle spasms and tremors
  - strained speech



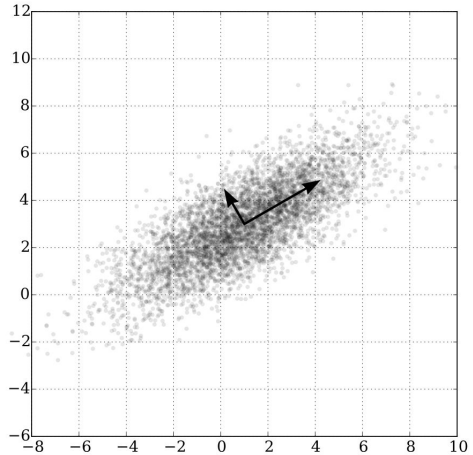
Features

- Fundamental Frequency (F0)
- Harmonic to Noise Ratio (HNR)
- Jitter (variation in Fundamental Frequency
- Shimmer (variation in Fundamental Amplitude)

[F0 is known to change with age, HNR and Jitter are known signs of degeneration, Jitter + Shimmer is known to be usable for distinguishing between speakers. Between all of them, this is a good starting point for building a model that can potentially detect degeneration beyond what is expected from aging alone]

Principal Component Analysis

- Principal Component Analysis is useful for removing redundancy from the data-frame and reducing the number of total variables
- Feature Extraction for Jitter and Shimmer yielded many related features
- Principal Component Analysis was able to remove this redundancy, and keep the feature count low



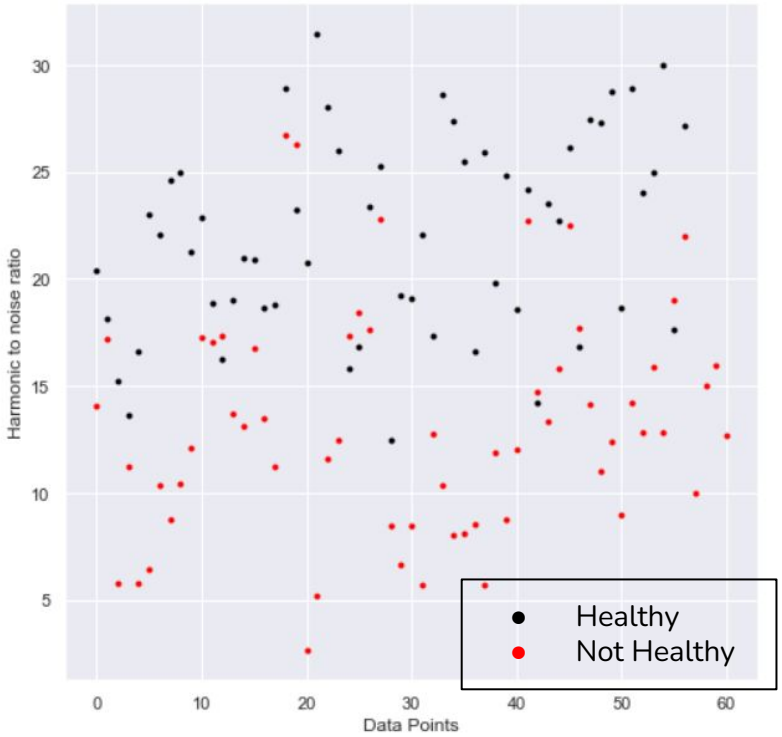
Linear Regression

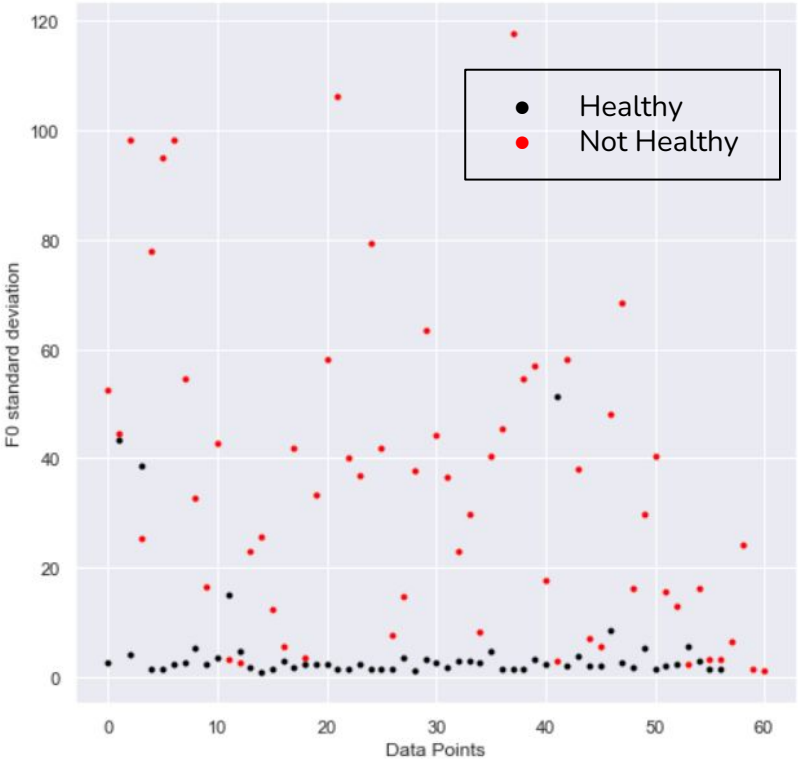
Linear Regression is a process for modeling the relationship between one or more independent variables and a single dependent variable.

Gradient Descent can be used to iteratively find the coefficients for a multi-term linear equation that minimize distance between predictions and actual results

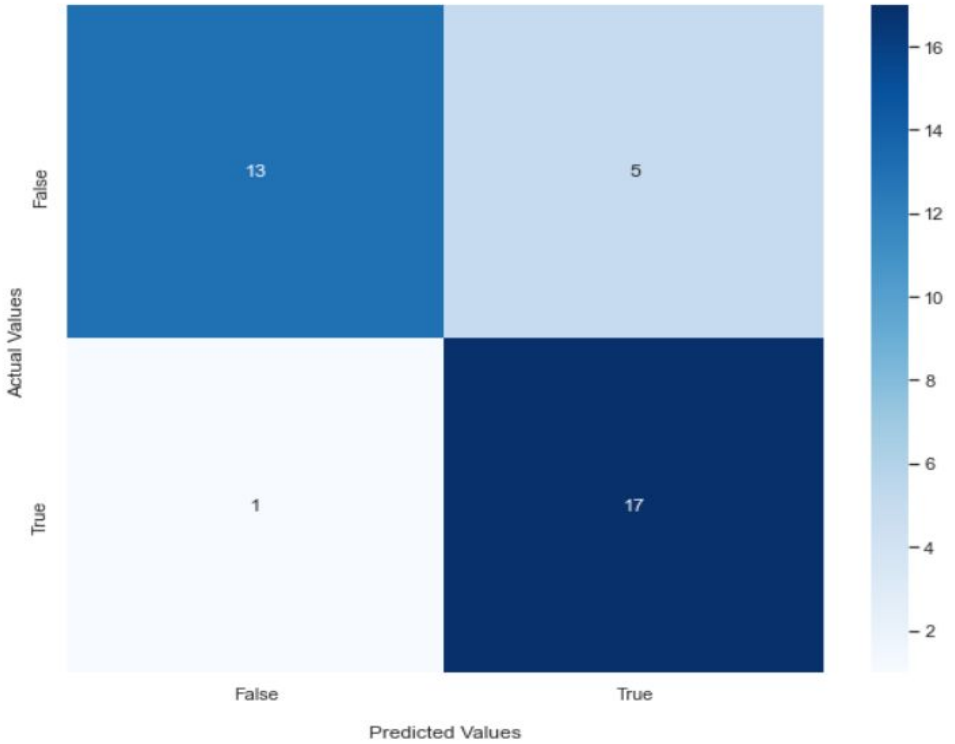
$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n$$

Findings


Harmonic to noise ratio
Data Points
Healthy
Not Healthy


F0 standard deviation
Data Points
Healthy
Not Healthy

Evaluation



Sensitivity = 94%

Specificity = 72%

Precision = 77%

Accuracy = 83%

Future Directions

Despite the accuracy of the current model, some improvements could be made to future improve accuracy or model utility:

- Inclusion of additional data sources, including ones that do not require just sustained vowel sounds
- Inclusion of additional meta-data (age, gender, etc) could improve ability to detect issues as these are also sources of variation in f0, jitter, and shimmer
- Expansion to differentiate between different disorders

Bibliography

Farrús, Mireia & Hernando, Javier & Ejarque, Pascual. (2007). Jitter and shimmer measurements for speaker recognition. Proceedings of the Interspeech 2007. 778-781. 10.21437/Interspeech.2007-147.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Murry T, Brown WS Jr, Morris RJ. Patterns of fundamental frequency for three types of voice samples. J Voice. 1995 Sep;9(3):282-9. doi: 10.1016/s0892-1997(05)80235-8. PMID: 8541971.

Nicoguaro, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons

Moore, M., Papreja, P., Saxon, M., et al. (2020). UncommonVoice: A Crowdsourced Dataset of Dysphonic Speech. *Proc. Interspeech.* 2532--2536