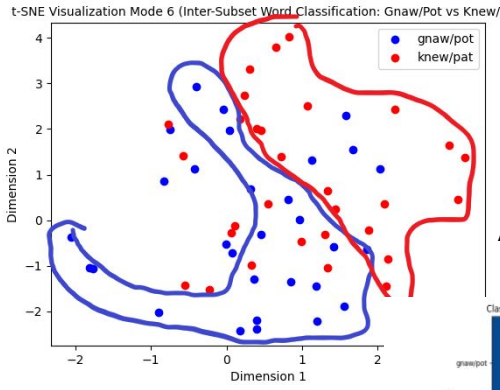
## Engineering Problem and Objectives

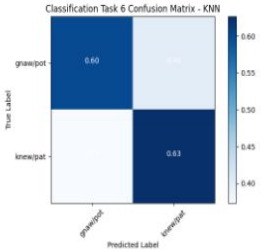
- 2 million people suffer from neuromuscular disorders - ALS, brain stem strokes, cerebral palsy, etc.
- Brain-computer interfaces (BCI's) have emerged as a potential solution
- BCI's have been built for communication, but they lack practicality and have a limited word-per-minute rate.
- Imagined/inner speech has emerged as a new potential paradigm
- This project seeks to determine the viability of decoding imagined speech from electroencephalography (EEG) signals through a variety of experimental classification modes.

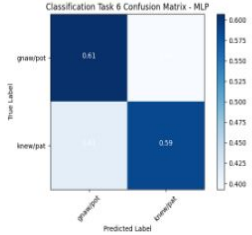
## Data Analysis and Results

- Encouraging results, including t-SNE visualization, testing accuracy, F1-score, and confusion matrix are shown for experimental/classification mode 6 (gnaw/pot vs pat/knew):




| | Testing Accuracy (mean +/- std.) | F1-score (mean +/- std.) |
|---|---|---|
| KNN | 61.33% (+/- 14.31%) | 60.21% (+/- 14.99%) |
| ANN | 59.67% (+/- 13.29%) | 57.49% (+/- 15.17%) |




## Project Design

Successful BCI construction consists of:

- **Dataset Collection**

  Online, publicly-available KARA-ONE dataset was used
  Contains EEG data corresponding to a variety of linguistics.

- **Data Preprocessing**

  Data preprocessed via passband + notch filter to remove unwanted and irrelevant artifacts.

- **Feature Extraction**

- Initially, 240 scalar-valued features extracted. Due to computational concerns, this was reduced to n = 30 features via Analysis of Variance (ANOVA), which ranks relevance of features depending on correlation to output classes.

- **Data Visualization**

- Inter-class discriminability and feature effectiveness was examined via t-Distributed-Stochastic-Neighbor-Embeddings (t-SNE), which maps high-dimensional spaces to a low-dimensional embedding for visualization

- **Classification**

- Used a baseline K-Nearest-Neighbor classifier and an experimental Artificial Neural Network classifier

  Optimized hyperparameters with Grid Search.

## Interpretations and Conclusions

- t-SNE visualization displayed relatively clear separation between classes (approximate clusters annotated on image)
- Mean testing accuracy was higher than chance level (chance = 50%).
- KNN yielded better performance over ANN

For all classification modes:

- Mean testing accuracy was greater than chance!
- EEG signals may indeed inherently encode information relevant to imagined speech.
- Future work includes experimentation with different paradigms (handwriting, images), and development of a fully functional BCI