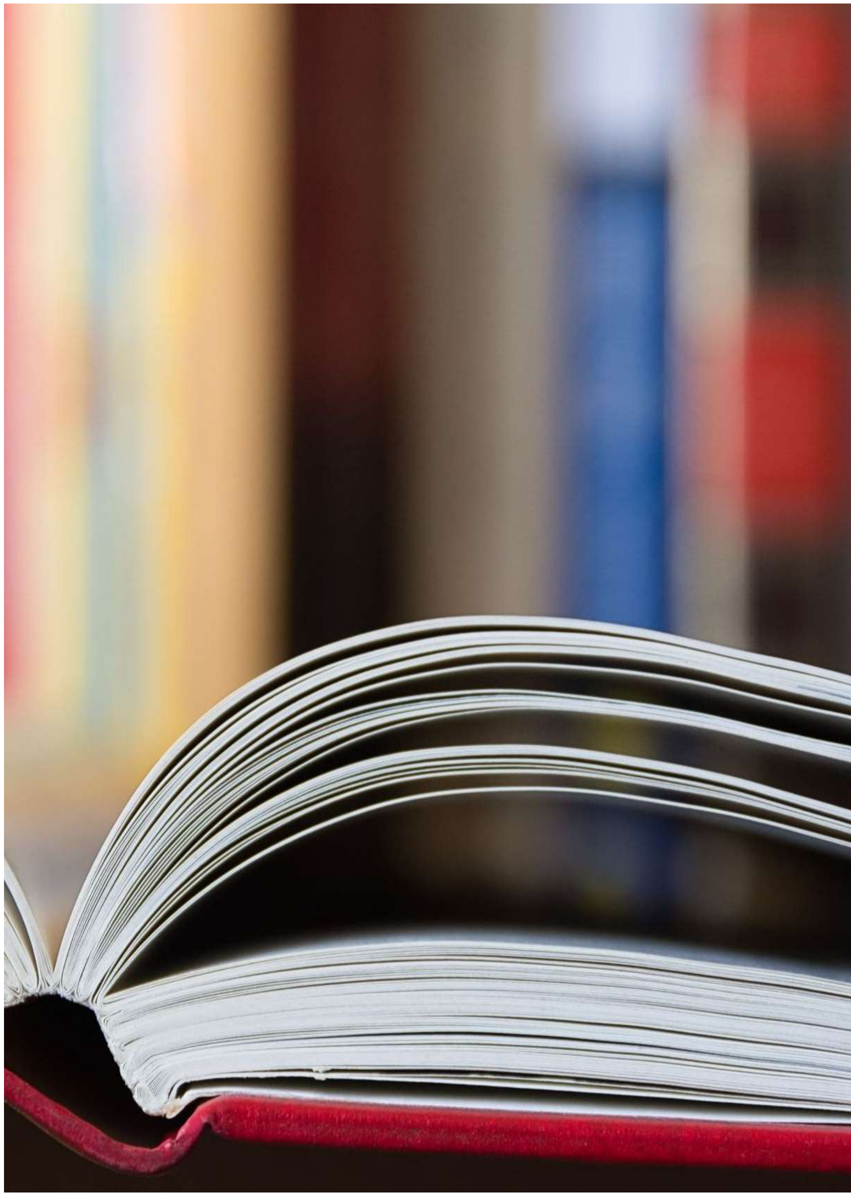
Tinnitus Data Classification Using Selected EEG Signal Connectivity Features

Presented by Penghui Zhao



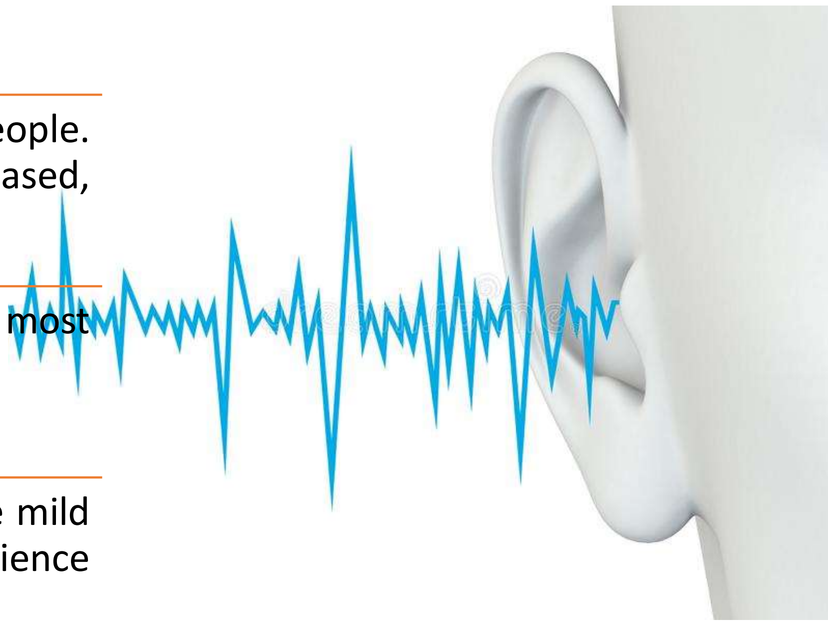
Background

Concepts: Tinnitus is a common disease in life. It is a hearing disorder that involves the perception of sound without an external source of sound.

Influence: It can cause distress, anxiety, and depression in people. In recent years, the prevalence of tinnitus has gradually increased, becoming one of the most pressing public health problems.

Reason: Hearing loss due to exposure to loud sound is the most common cause of tinnitus.

NZ situation: Up to 30% of people in New Zealand experience mild tinnitus at some point in their lives, and around 5% experience painful tinnitus.



Motivations

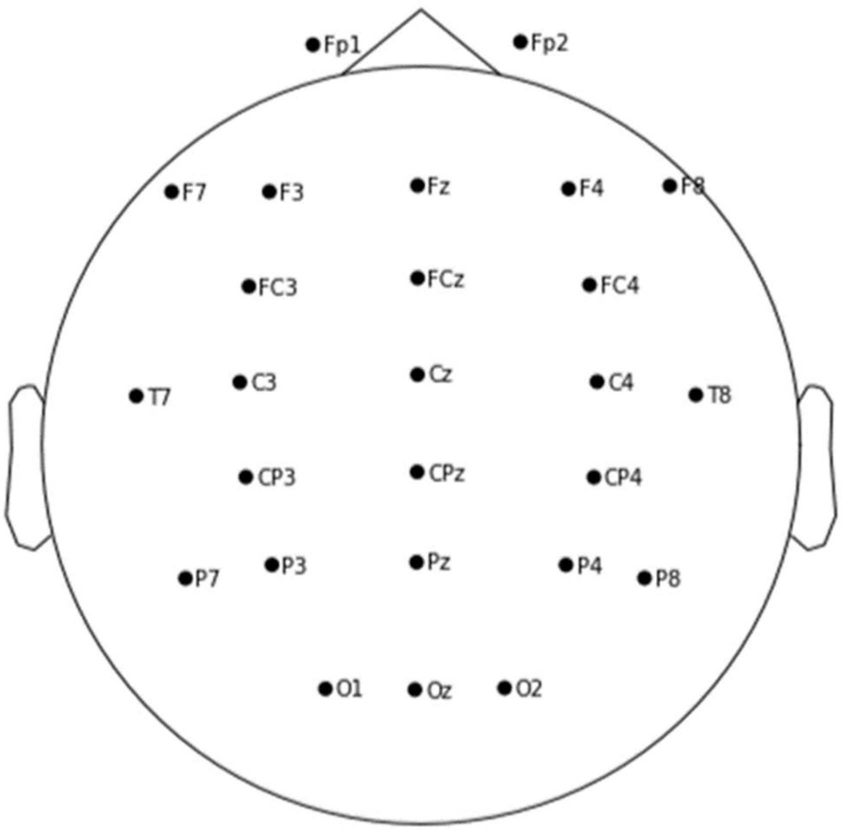
- **Problem:** Tinnitus is a subjective symptom or feeling; there is still a lack of an objective method to diagnose and detect tinnitus symptoms.

- **What we can do:** This study aims to use machine learning algorithms to classify tinnitus data, select the best classifier, and apply it in the clinical detection of tinnitus. This study would help doctors improve the efficiency of tinnitus detection so that tinnitus patients could get faster and more effective treatment in the later stages.

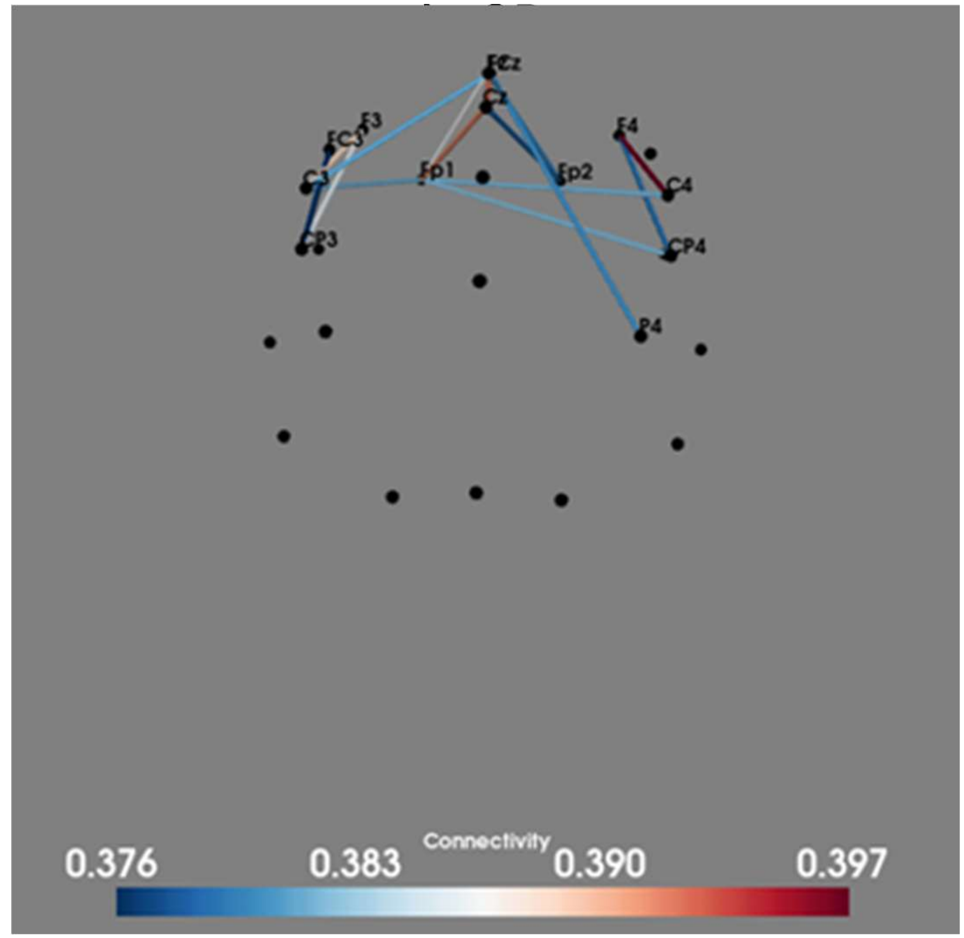
Experiment related concepts

- **The EEG (Electroencephalogram) signal** indicates the electrical activity of the brain. It is a non-invasive technique used to measure multi-channel potentials.
- **Epochs** are specific time-windows extracted from the continuous EEG signal, usually time-locked with respect to an event.
- **Power Spectral Density** shows the intensity of the frequency components of an EEG signal, indicating at which frequencies the variance of the data is large.
- **Functional connectivity** is a way to measure the degree of correlation between every pair of EEG channels.
- **Principal component analysis** (PCA) is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.
- **The feature importance attribute** of the model can be used to obtain the feature importance of each feature in the dataset. Feature importance assigns a score to each of the data's features; the higher the score, the more important or relevant the feature is to the output variable.

26 Channels used in experiment



Functional connectivity example



Literature review

In these papers, the number of features involved is limited, so the results may not be universal. The experimental design steps and ideas in these papers are similar, but the amount of data and feature selection methods are different, which will affect the experimental results.

Researchers	Main Work	Classifiers used	Features used	Prediction Results
(Sun et al., 2019)	Classify the multi-view features of tinnitus patients and healthy individuals	SVM	Multi-view features in EEG signals	Accuracy was 99.23%
(Doborjeh et al., 2023)	Predict the results of tinnitus treatment	Deep learning	Frequency and functional connectivity Features	Accuracy was 99%.
(Li et al., 2022)	Diagnose the location of tinnitus	SVM, MLP (multilayer perceptron)	Connectivity and time-frequency domain features	Average accuracy of the SVM and MLP reached 99.4% and 99.1%, respectively
(Fahimeh et al., 2019)	Distinguish healthy and tinnitus subjects automatically	SVM	Graph-theory-based metrics extracted from EEG data	Accuracy was 100% in the Beta2 frequency band

- Can we use a large number of functional connectivity features to do the tinnitus data classification?

- What feature dimension reduction method can we use to improve the results?

- What machine learning classifiers can we use? Which classifier works best?

Research questions

Data and methods

- The data came from the neurophysiological EEG database. It is a resting state EEG signal dataset that includes 33 Tinnitus patients and 47 Healthy individuals' EEG information.
- The initial dataset was divided into a 90% training dataset used for internal testing and a 10% testing dataset used for external testing.
- Use the Python MNE package to pre-process the data: remove the bad channels filter the frequency range, segment the 120-seconds data into 6 epochs, extract the Power Spectral Density (PSD), compute the functional connectivity features, and finally obtain 325 features.
- Feature dimension reduction methods such as Principal Component Analysis (PCA) and Python Feature Importance were used. These two methods reduced the full 325 features to 200, 150, 100, 50, and 20 features respectively.

Data and methods

- Balance the training dataset to avoid the prediction results being biased for each subset of features.
- Ten-fold cross-validation as the validation method was adopted to measure the performance of each model on the training data and find the best hyperparameters for each model.
- Final classification models on different subset of features were built by using the best hyperparameters for each model that found by the cross validation. The performance of external testing on the testing data was measured to test the generalization of each final models.
- The performance evaluation method was Accuracy. The classification accuracy rate is the ratio of the number of samples correctly classified by the classifier divided by the total number of samples being predicted.

Results: Classification accuracy rate on training dataset

- The figure on the right side shows the Classification accuracy results on training dataset. The five classifiers accuracy rates under different numbers of feature look different.
- Regarding internal testing on the training dataset , we found that logistic regression had the best accuracy rate of 100% on all the different numbers of selected features. The second-best classifier was SVM, which also had a 100% accuracy rate on most of the subsets of features.
- Based on the internal testing, we can get the best hyperparameters which can be used for the final classification models.

Connectivity Features	Number of Features	LR	KNN	DT	SVM	RF
Full features	325	1	0.68	0.75	1	0.875
PCA feature extraction	200	1	0.68	0.60	1	0.83
PCA feature extraction	150	1	0.68	0.59	1	0.81
PCA feature extraction	100	1	0.68	0.62	1	0.83
PCA feature extraction	50	1	0.68	0.68	1	0.85
PCA feature extraction	20	1	0.66	0.67	0.90	0.78
Python feature importance	200	1	0.73	0.74	1	0.86
Python feature importance	150	1	0.73	0.76	1	0.89
Python feature importance	100	1	0.74	0.77	1	0.93
Python feature importance	50	1	0.75	0.85	1	0.86
Python feature importance	20	1	0.72	0.71	1	0.85

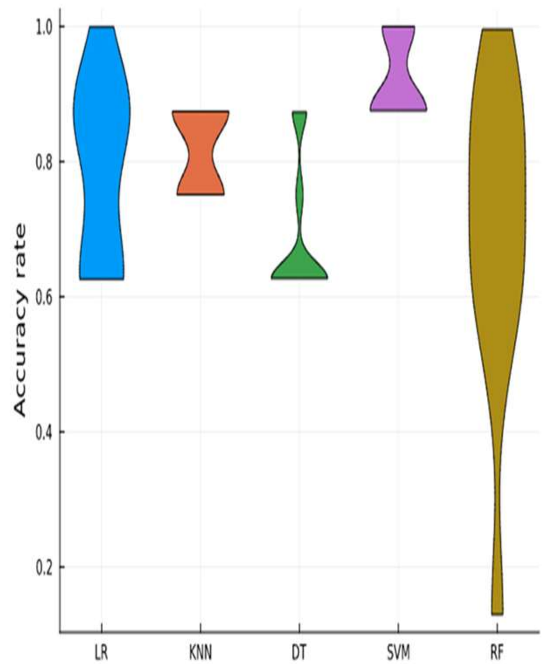
Results: Classification accuracy rate on testing dataset

- Regarding the external test on the testing dataset, we found SVM had the best accuracy rate; the average accuracy is 92%. Logistic Regression, KNN, Decision Tree, and Random Forest achieved average accuracy of 78%, 82%, 68%, and 69%, respectively.
- The SVM classifier with 20 features that were extracted by the Python feature deduction method should be selected as the best classifier in this experiment as it had the highest classification accuracy rate and lowest number of features.

Connectivity Features	Number of Features	LR	KNN	DT	SVM	RF
Full features	325	0.875	0.75	0.625	0.875	0.5
PCA feature extraction	200	0.625	0.75	0.625	0.875	0.875
PCA feature extraction	150	0.625	0.75	0.625	0.875	0.625
PCA feature extraction	100	0.625	0.75	0.625	0.875	0.75
PCA feature extraction	50	0.875	0.875	0.625	0.875	0.625
PCA feature extraction	20	0.625	0.875	0.875	0.875	0.125
Python feature importance	200	0.875	0.875	0.625	1	0.875
Python feature importance	150	1	0.875	0.75	1	0.625
Python feature importance	100	0.875	0.75	0.625	0.875	0.75
Python feature importance	50	0.875	0.875	0.625	1	1
Python feature importance	20	0.75	0.875	0.875	1	0.875

Accuracy results comparison on the testing dataset

- We can also use statistical tests such as the T-test and Mann-Whitney U-test to compare the performance of each two classifiers. These two statistical tests were employed to test whether the means of the accuracy of each two classifiers are statistically the same, this is the null hypnosis for these two statistical tests.
- For example, when test SVM and Random Forest classifiers, both test summary showed reject hypnosis null. That means there is at least 95% confidence that the classification accuracy rates of SVM classifier and Random Forest classifier are statistically different.
- Then the violin plot needs to be employed to check which classifier accuracy had the higher accuracy rate since the statistical tests only shows their means are different. Violin plot shows different accuracy density. From the violin plot can see SVM was better than Random Forest as the majority accuracy (widest part) of SVM layed around just under 87.5%, whereas Random Forest majority accuracy is around 70%.




```
Two sample t-test (unequal variance)
------------------------------------
Population details:
    parameter of interest:   Mean difference
    value under h_0:         0
    point estimate:          0.227273
    95% confidence interval: (0.06355, 0.391)

Test summary:
    outcome with 95% confidence: reject h_0
    two-sided p-value:           0.0108

Details:
    number of observations:   [11,11]
    t-statistic:              3.042903097250921
    degrees of freedom:       11.379512195121952
    empirical standard error: 0.0746894396597954
```

```
Approximate Mann-Whitney U test
-------------------------------
Population details:
    parameter of interest:   Location parameter (pseudomedian)
    value under h_0:         0
    point estimate:          0.125

Test summary:
    outcome with 95% confidence: reject h_0
    two-sided p-value:           0.0049

Details:
    number of observations in each group: [11, 11]
    Mann-Whitney-U statistic:              101.5
    rank sums:                             [167.5, 85.5]
    adjustment for ties:                   1140.0
    normal approximation (μ, σ):           (41.0, 14.3887)
```

Conclusion

- We can use 325 functional connectivity features extracted from the EEG signal for tinnitus data classification.
- The feature dimension reduction methods PCA and Python Feature Importance can be used to reduce the 325 features to 200, 150, 100, 50, and 20 separately. From the classification accuracy results, we can see that Python Feature Importance is better.
- The five machine learning classifiers (Logistic Regression, KNN, Decision Tree, SVM, Random Forest) can be used for the tinnitus data classification.
- The best machine learning classifier is Support Vector Machine (SVM), which achieved an average classification accuracy of 92% for all subsets of features. The highest accuracy for SVM was 100% with the 20 sub-features.

Future work

There are still some potential works to be done in the future to better improve the experimental results. For example:

Split the data into more epochs to explore how the experimental results differ.

Reduce the number of features to different numbers, such as 120,60,30,10 to see any impact on the prediction results.

Use other dimensionality reduction methods besides PCA and Python feature importance.

Adopt deep learning algorithm to the experiment which may lead to some new discoveries.

Thank you!