MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
*(A constituent unit of MAHE, Manipal)*

# Enhancing EEG-based Imagined Speech Recognition through Spatio-Temporal feature extraction using Information Set Theory

## Final Semester Thesis Presentation

**Ashrith Sagar Yedlapalli**
**200902016**

*Under the guidance of*
**Dr. Jeevan M**

**Department of Biomedical Engineering,**
**Manipal Institute of Technology, Manipal, India**

July 6, 2024

# Outline I

# Outline II

# Introduction

## Area of work

### Electroencephalography (**EEG**) and **Imagined/Silent Speech** (SS)

- **EEG** $\rightarrow$ A non-invasive brain activity recording method.
- **SS** $\rightarrow$ Speech silently formed in the mind *without verbal articulation*.
- **Advantages** of EEG-based BCI:
  - **Non-invasive** and easy to use.
  - **Good temporal resolution**.
- **Challenges** in SS recognition using EEG:
  - Low Signal-to-Noise Ratio (SNR).
  - Limited spectral and spatial resolution.

# Introduction

## Motivation

### Rationale behind the work

- EEG signals are integral to Brain-Computer Interfaces (BCI), enhancing interaction for paralyzed individuals [11].
- BCIs translate brain signals into **actionable commands**, *bypassing natural neuromuscular pathways*.
- Existing BCI paradigms are limited by specific stimuli and class options, hindering practical communication [5].
- EEG signals are utilized in recognizing Imagined Speech (Covert or Silent speech), where users mentally visualize words instead of speaking [5].
- EEG's **non-invasive nature** and **ease of use** make it ideal for recognizing imagined speech.

# Introduction

## Objectives

### Project Objectives

- Conduct a EEG Imagined speech **phonological classification**
- Extracting **rich spatio-temporal features** using Information Set theory

# Review

EEG

## Electroencephalography (**EEG**)

- EEG is a **non-invasive** method to record brain electrical activity using scalp electrodes [6].
- Since Hans Berger first recorded EEG signals in 1924, EEG has been pivotal in neuroscience research, providing insights into brain activity [14].
- **EEG waveforms** are categorized into Delta (0.5 to 4 Hz), Theta (4 to 8 Hz), Alpha (8 to 12 Hz), Beta (12 to 35 Hz), and Gamma (greater than 36 Hz) bands [7].
- **Preprocessing** involves downsampling, filtering, and windowing to enhance signal quality [7].
- **Feature extraction** analyzes time, frequency, and spatial domains to capture meaningful data for classification [7].

# Review

Imagined Speech

## Imagined/Silent Speech (SS)

- Imagined speech, or covert speech, involves mentally representing words or concepts without verbal articulation [5].
- for imagined speech applications generally comprises four steps [7]
  - Signal acquisition
  - Signal processing
  - Feature extraction
  - Classification
- Classification methods include traditional ML (e.g., SVM, kNN, RF) and deep learning (e.g., CNN, LSTM) [7].
- Deep learning models like CNN and LSTM show high performance but require large training data and are computationally expensive.
- Limited data

# Review

## KaraOne database

### KaraOne database [13]

- A publicly available dataset containing data from **3 modalities**: **Acoustic**, **Facial**, and **EEG data**.
- Contains EEG signals from **14 subjects**, recorded during imagined speech tasks.
- Each subject performed 15 trials, each consisting of **four states**: *Rest*, *Stimulus*, *Imagined speech*, and *Speaking*.

# Review

## KaraOne database — Recording

### Recording procedure

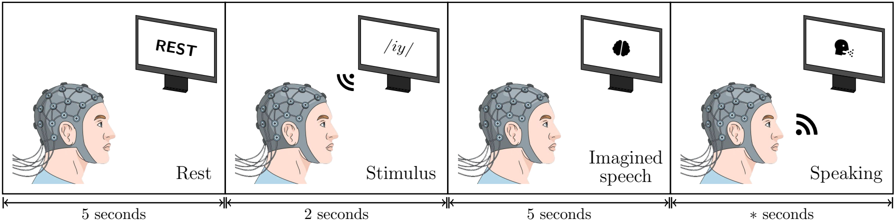


Figure: A visual representation of the recording procedure used in the KaraOne database. Each trial for each subject consisted of **four states**: *rest*, *stimulus*, *imagined speech*, and *speaking*. The imagined speech segment is the focus of this study. [Source: Author; Refer [13]]

# Review

KaraOne database — Prompts

## Prompts

- The cues consist of

| | |
|---|---|
| **Seven** phonemic/ syllabic prompts: | */diy/*, */iy/*, */m/*, */n/*, */piy/*, */tiy/*, */uw/* |
| **Four** words from Kent's phonetically-similar pairs: | gnaw, knew, pat, pot |

- They were chosen in a way to ensure a **balanced representation** of nasals, plosives, vowels, voiced, and unvoiced phonemes.

# Review



## FEIS dataset

- Acquisition System: **14-channel Emotiv EPOC+ headset** with dry electrodes
- Sampling frequency: 256 Hz
- Dataset contents:
  - **16 phonemic prompts**
  - 10 repetitions of each prompt per participant
  - 21 participants in total
- Signal pre-processing:
  - Notch filter applied at 50 Hz and 60 Hz using built-in Emotiv software to remove powerline noise
  - No pre-processing for physiological artifacts (blinks, saccades) due to *lack of ocular channels* in the headset
- Epoch Segmentation:
  - EEG epochs segmented into 5-second intervals
  - 10% overlap between consecutive epochs
- Similar preprocessing steps to the KaraOne dataset

# Review

## FEIS dataset

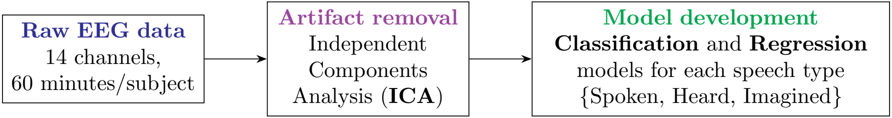


Figure: Illustration of the model-building procedure for the FEIS dataset.

# Review

## FEIS dataset

Table: Comparison of KaraOne and FEIS datasets

| | **KaraOne** [13] | **FEIS** [12] |
|---|---|---|
| **EEG device** | 64 channels | 14 channels |
| **Sampling rate** | 1000 Hz | 256 Hz |
| **Participants** | 14 participants | 21 participants |
| **Duration** | 30 - 40 minutes | 60 minutes |
| **Prompts** | 7 phonemes + 4 words | 16 phonemes |

# Review

Information set theory

## Information set theory

- **Information set theory** is a **mathematical framework** that deals with the representation and manipulation of *information sets*.
- **Information sets** are **mathematical objects** that represent the **information content** of a given *fuzzy set*.

Information set theory is used in this study to **extract spatio-temporal features** from EEG signals.

# Review

Information set theory

## Entropy functions

- **Entropy functions** are used to **quantify the uncertainty** and **measure the information content** in a given *information set*.

Table: Entropy functions

| Entropy | Formula |
|---|---|
| Shannon entropy | $E_{Shannon} = -\sum_i p(x_i) \log p(x_i)$ |
| Shannon fuzzy entropy | $E_{X,LT,\mu} = -\frac{1}{n} \sum_i (\mu_X^k(x_i)) \log (\mu_X^k(x_i))$ $+(1-\mu_X^k(x_i)) \log (1-\mu_X^k(x_i))$ |

# Review

Information set theory

## Information sets

- Applications in extracting effective information from fuzzy sets.
- Utilized in studies by Aggarwal et al. (2016), Medikonda et al. (2020), Mamta et al. (2014).
- Information sets are **derived from fuzzy sets**; Values from the fuzzy set are treated as information source values.
- Source values, upon multiplication with an entropy function yields information values.
- The **sum** of information values in an information set indicates **information**/uncertainty.

# Review

## Information set theory

### Information sets

- Solves two main problems involved with fuzzy sets:
  - **Decouples** information source values and membership grades in fuzzy sets.
  - **Extends** entropy function to probabilistic or possibilistic domains.

# Methodology

## Overview

- **Data pre-processing**
  - Filtering & Baseline correction
- **Feature extraction**
  - Windowing
  - Extracting features
  - Additional features
  - Feature reduction
    - KaraOne methodology
    - Proposed IFS methodology
- **Classification**
  - Model training & evaluation

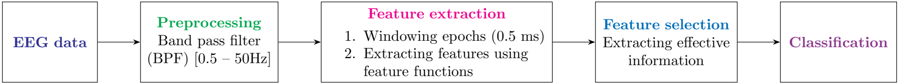


Figure: Methodology flowchart for the proposed work.

# Methodology

Classification tasks

## Classification tasks

**Five** **binary** classification tasks of *phonological categories* are performed

1. Vowel-only vs. consonant (**C/V**)
2. Presence of nasal ($\pm$ **Nasal**)
3. Presence of bilabial ($\pm$ **Bilabial**)
4. Presence of high-front vowel ($\pm$**/iy/**)
5. Presence of high-back vowel ($\pm$**/uw/**)

# Methodology

Pre-processing

## Filtering

- The data was filtered using a **band-pass filter** with
  - Lower cut-off frequency of 0.5 Hz
  - Upper cut-off frequency of 50 Hz.
- This was done to **remove the noise** from the data.

## Baseline correction

- The data was **baseline corrected** using the mean of the first 500 ms of each trial.
- This was done to **remove the DC offset** from the data.

The data tensor was then used for further processing.

# Methodology

## Pre-processing

### Windowing

- Windowing was performed over each imagined speech segment of 5 seconds.
- This gave 19 windows per segment, with 500 ms window length.

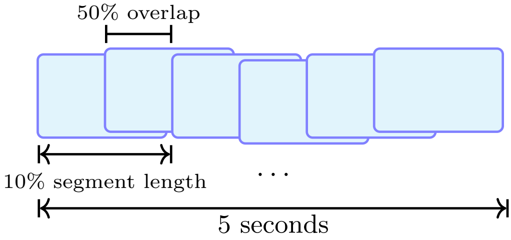


Figure: Windowing of the data

# Methodology

## Pre-processing

### Feature extraction

- A set of statistical features and few entropy functions were computed on the windows, giving a total of 27 features per window.
- This results in a feature tensor of size $19 \times 62 \times 27$.

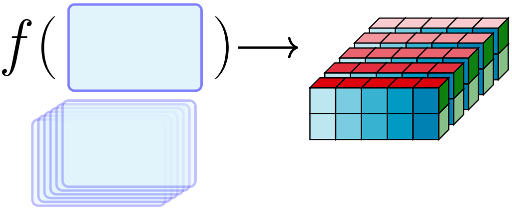


Figure: Extraction of features on the windows

# Methodology

## Pre-processing

Table: List of the 27 functions used for feature extraction on the windows. The functions are divided into three categories: **statistical measures**, **entropies**, and **fractal dimensions**.

| Feature functions | |
|---|---|
| Mean | Absolute Mean |
| Maximum | Absolute Maximum |
| Minimum | Absolute Minimum |
| Minimum $\pm$ Maximum | Curve Length |
| Energy | Nonlinear Energy |
| Integral | Standard Deviation |
| Variance | Skewness |
| Kurtosis | Sum |
| Spectral entropy | Sample entropy |
| Permutation entropy | Singular Value Decomposition (SVD) entropy |
| Approximate entropy | Petrosian fractal dimension |
| Katz fractal dimension | Higuchi fractal dimension |
| Root Mean Square | Detrended fluctuation |

# Methodology

## Pre-processing

### Additional features

- **Delta** and **double delta** features were computed on the 27 features, giving a total of 81 features per window.
- This results in a feature tensor of size $17 \times 62 \times 81$.
- The initial two windows were discarded to accommodate the inclusion of these additional features.

# Methodology

## Pre-processing

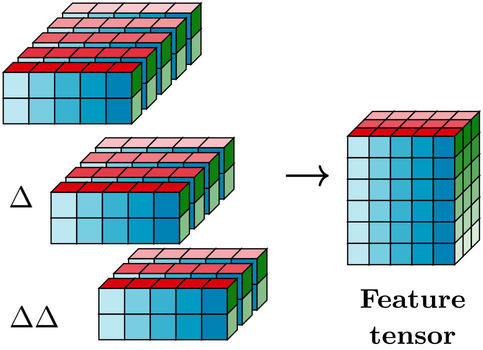


Figure: Addition of **delta** and **double delta** features to the feature tensor. The initial two windows were discarded to accommodate the inclusion of these additional features.

# Methodology

## Proposed methodology

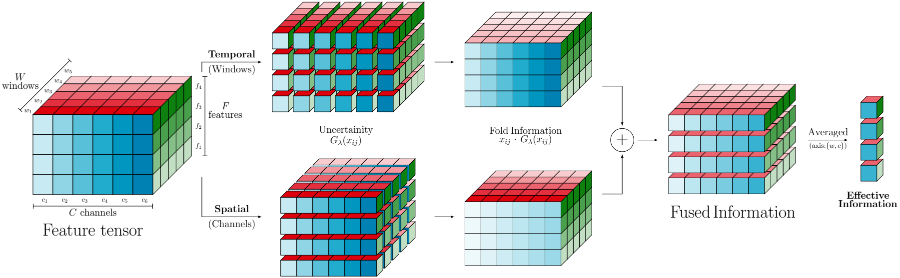


Figure: Extracting effective information from the features across both temporal and spatial dimensions resulting in rich **spatio-temporal features**. The two folds of information are separately computed across the temporal and spatial dimensions by extracting the uncertainty in the source values. These are then fused and averaged along the temporal and spatial dimensions to obtain the effective feature vectors.

# Methodology

Proposed methodology

**Extracting Effective information — Information Sets**

The procedure for computing effective features from the feature tensor is as follows:

# Methodology

## Proposed methodology

For each epoch, consider the feature tensor $X$ of size $[W \times C \times F]$.

Consider the feature matrix $X_f$ for the $f^{th}$ feature in (1), which is considered as an **information source matrix**. The features are considered as **information source values** comprising an information set along the temporal and spatial dimensions.

$$X_f = \begin{bmatrix} x_{11f} & \cdots & x_{1cf} & \cdots & x_{1Cf} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{w1f} & \cdots & x_{wcf} & \cdots & x_{wCf} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{W1f} & \cdots & x_{Wcf} & \cdots & x_{WCf} \end{bmatrix}_{W \times C} \qquad \begin{matrix} 1 \leq w \leq W \\ 1 \leq c \leq C \\ 1 \leq f \leq F \end{matrix} \tag{1}$$

# Methodology

## Proposed methodology

The uncertainty across the temporal and spatial dimensions are computed separately using a Gaussian membership function in (2), (3) as follows:

$$G_t\left(x_{wcf}\right) = \exp\left\{-\frac{1}{2}\left[\frac{x_{wcf} - \mu_{cf}}{\sigma_{cf}}\right]^2\right\}, \quad \begin{array}{c} 1 \leq c \leq C \\ 1 \leq f \leq F \end{array} \tag{2}$$

$$G_s\left(x_{wcf}\right) = \exp\left\{-\frac{1}{2}\left[\frac{x_{wcf} - \mu_{wf}}{\sigma_{wf}}\right]^2\right\}, \quad \begin{array}{c} 1 \leq w \leq W \\ 1 \leq f \leq F \end{array} \tag{3}$$

# Methodology

## Proposed methodology

Here $\mu_{cf}$ and $\sigma_{cf}$ are the temporal mean and standard deviation calculated across the windows in (4), while $\mu_{wf}$ and $\sigma_{wf}$ are the spatial mean and standard deviation calculated across the channels in (5).

$$\mu_{cf} = \frac{1}{W} \sum_{w=1}^{W} x_{wcf}, \quad \sigma_{cf} = \sqrt{\frac{1}{W} \sum_{w=1}^{W} (x_{wcf} - \mu_{cf})^2} \tag{4}$$

$$\mu_{wf} = \frac{1}{C} \sum_{c=1}^{C} x_{wcf}, \quad \sigma_{wf} = \sqrt{\frac{1}{C} \sum_{c=1}^{C} (x_{wcf} - \mu_{wf})^2} \tag{5}$$

# Methodology

## Proposed methodology

The temporal and spatial fold information in (6) are then obtained by extracting information from the uncertainties in the source values, thereby considering the set $\left\{ I_f^{\lambda}\left(x_{wcf}\right)\right\}$ as an information set, which is given by,

$$I_f^{\lambda}\left(x_{wcf}\right) = x_{wcf} \cdot G_{\lambda}\left(x_{wcf}\right), \quad \lambda = \left\{t, s\right\} \tag{6}$$

where $\lambda$ corresponds to the folds, representing the temporal $(t)$ and spatial $(s)$ folds respectively.

# Methodology

## Proposed methodology

The choice of the **membership function** as a **Gaussian** in (6) is based on the fact that the **mean** of the attributes should remain the same along the temporal and spatial dimensions in the feature matrix, taken independently.

This is justified based on the assumption that the signals remain **statistically stationary** over time within the windows, and that the signals are **localised** within their respective spatial regions for each channel.

# Methodology

## Proposed methodology

The fused information for a given window $w$ and channel $c$ is then computed by summing the fold information across the corresponding window and channel in (7).

$$\mathcal{I}_f\left(x_{wcf}\right) = I_f^t\left(x_{wcf}\right) + I_f^s\left(x_{wcf}\right), \quad \begin{array}{c} 1 \leq w \leq W \\ 1 \leq c \leq C \\ 1 \leq f \leq F \end{array} \tag{7}$$

# Methodology

## Proposed methodology

The normalised effective information is then computed by taking the mean at each fused information source value along the temporal and spatial folds in (8).

$$E_f = \frac{1}{WC} \sum_{w=1}^{W} \sum_{c=1}^{C} \mathcal{I}_f \left( x_{wcf} \right) \tag{8}$$

This gives the following $F \times 1$ effective information vector $\mathbf{E}$,

$$\mathbf{E}^T = \begin{bmatrix} E_1 & \cdots & E_f & \cdots & E_F \end{bmatrix}_{1 \times F} \tag{9}$$

These are the effective features that are used to represent the spatio-temporal information, thus effectively reducing the $[W \times C \times F]$ feature tensor to a $F$-length feature vector, as in (9).

# Methodology

## Proposed methodology

**Algorithm 1:** Effective Information

**Input:** $[S \times W \times C \times F]$ feature tensor with $S$ epochs, $W$ windows, $C$ channels, and $F$ features

**foreach** *epoch in feature tensor* **do**

Compute the temporal and spatial fold informations

$$G_t\left(x_{wcf}\right) = \exp\left\{-\frac{1}{2}\left[\frac{x_{wcf}-\mu_{cf}}{\sigma_{cf}}\right]^2\right\},\ G_s\left(x_{wcf}\right) = \exp\left\{-\frac{1}{2}\left[\frac{x_{wcf}-\mu_{wf}}{\sigma_{wf}}\right]^2\right\},$$

$$I_f^{\lambda}\left(x_{wcf}\right) = x_{wcf} \cdot G_{\lambda}\left(x_{wcf}\right), \lambda = \left\{t, s\right\}$$

Compute the fused information from the fold informations

$$\mathcal{I}_f\left(x_{wcf}\right) = I_f^t\left(x_{wcf}\right) + I_f^s\left(x_{wcf}\right)$$

Compute the effective information vector $\mathbf{E}$

$$E_f = \frac{1}{WC}\sum_{w=1}^{W}\sum_{c=1}^{C}\mathcal{I}_f\left(x_{wcf}\right)$$

**Output:** Feature vector of length $[F]$ for the epoch

**end**

**Result:** Effective feature matrix of dimensions $[S \times F]$

# Methodology

## KaraOne methodology

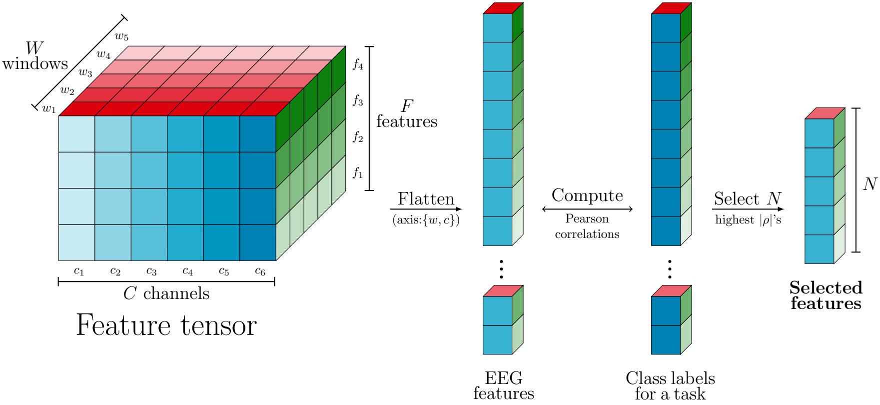


Figure: EEG features are **ranked** based on their Pearson correlations with the given classes for each task *independently*. $N$ features are selected with the highest absolute correlation coefficients, where $N \in \{5, \ldots, 100\}$. [Source: Author; Refer [13]]

# Methodology

Models

## Models

We use the following models to classify the EEG signals:

- **Hanman Classifier**
- **Decision Tree Classifier**
- **Random Forest Classifier**
- **Support Vector Machines**

# Methodology

Models — Hanman Classifier

## Hanman Classifier

- Based on Information sets [9, 8].
- Works by computing the uncertainty of the minimum aggregated normed errors between the test sample and all the training samples for each class and then identifies the class with the lowest uncertainty as the predicted class for the test sample.
- The aggregation of the errors between the test and training samples is done using the Frank t-norm to assess similarity or dissimilarity between them.

# Methodology

## Models — Hanman Classifier

**Algorithm 2:** Hanman Classifier

**Input:** Train samples, Test samples
Normalize samples along features axis
**foreach** *sample in test samples* **do**
    **foreach** *sample in train samples* **do**
        Compute the error between the training sample and the input test sample
    **foreach** *class* **do**
        **foreach** *pair of errors in class* **do**
            Compute the $t$-norm of the error vectors
        Compute the minimum of the normed error pair
        Compute the possibilistic uncertainty associated with the minimum pair
    Compute the class with the minimum uncertainty
    **Output:** Predicted class for the test sample
**Result:** Predicted classes for the test samples

# Methodology

## Models — Hanman Classifier

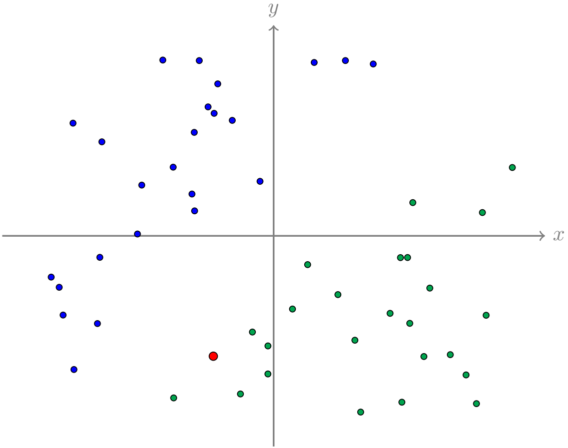


Figure: Setup of the data points in the classification problem.

# Methodology

## Models — Hanman Classifier

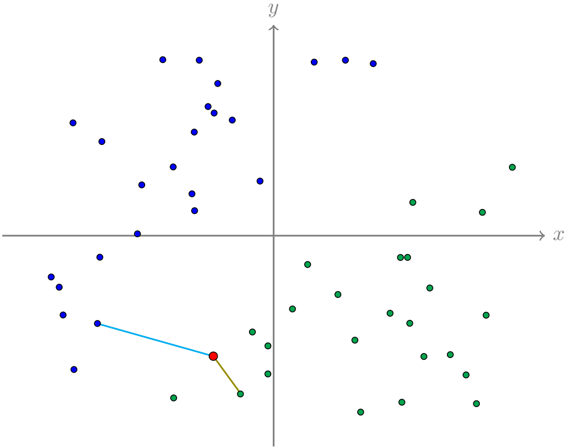


Figure: Hanman Classifier classifies by computing the uncertainty of the minimum aggregated normed errors between the test sample and all the training samples for each class and then identifies the class with the lowest uncertainty as the predicted class for the test sample.

# Methodology

Models — Decision Tree Classifier

## Decision Tree Classifier [10]

- Works by approximating the features and learning them by a series of if-then rules, recursively dividing data into subsets based on features that maximize information gain.
- Information gain is computed using either the Gini impurity function or the Shannon entropy function:

$$E_{Gini} = \sum_k p_k(1 - p_k) \qquad E_{Shannon} = -\sum_k p_k \log p_k$$

- Its complexity is closely related to its depth.
- Tend to overfit, do not generalize well, and are sensitive to changes in the data.

# Methodology

## Models — Random Forest Classifier

### Random Forest Classifier [2]

- An ensemble method that works by creating a forest of decision trees.
- The forest's outcome relies on a **majority-voting principle**, selecting the most predicted class (mode of the predictions) from the ensemble.
- By aggregating predictions of multiple trees, RFC achieves higher accuracy and **better generalization** compared to a single DT.
- Can handle large datasets and is robust to overfitting, overcoming the problems in individual decision trees.

# Methodology

## Models — Random Forest Classifier

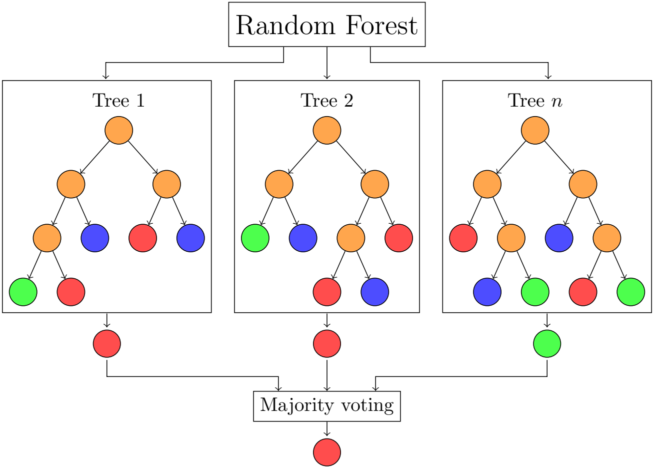


Figure: Random Forest Classifier: An ensemble of multiple Decision Trees.

# Methodology

Models — SVM

## Support Vector Machine (SVM)

- SVMs operate by identifying the **optimal hyperplane** to effectively separate data into distinct classes.
- This involves finding a hyperplane with the **maximum margin** between classes, thereby maximizing the separation between the hyperplane and the closest data points from each class.
- SVMs excel in high-dimensional areas and can achieve nonlinear classification by utilizing kernel functions that convert input data into feature spaces of higher dimensions.

# Methodology

## Models — SVM

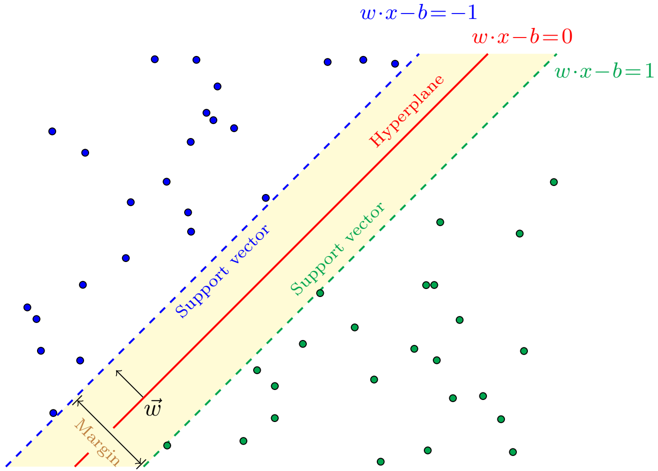


Figure: The classifier finds the hyperplane that best separates the data into different classes.

# Methodology

## Models — Parameters

### Model parameters

A **Grid search** was performed to obtain the best parameters for each model.

Table: Parameters for the different models

| Model | Parameters |
|---:|---|
| Hanman Classifier | `alpha=0.5, beta=1, a=1, b=0, q=2` |
| Decision Tree Classifier | `min_samples_split=2, max_depth=None, criterion='gini'` |
| Random Forest Classifier | `max_depth=100, n_estimators=100, class_weight='balanced'` |
| Support Vector Machines | `C=1, kernel='rbf', gamma='scale', shrinking=True` |

# Methodology

Samplers

## Samplers

- **Random Over Sampler** (ROS)
- **SMOTE** (Synthetic Minority Over-sampling Technique) [3]
- **ADASYN** (Adaptive Synthetic Sampling) [4]

# Methodology

## Samplers — Random Over Sampler

### **Random Over Sampler** (ROS)

- A simple method that addresses class imbalance by **randomly** selecting and replicating samples from the *minority class*.
- Over-samples by duplicating original minority class samples
- Need to be used with caution to avoid overfitting.

# Methodology

## Samplers — SMOTE

### Synthetic Minority Over-sampling Technique (**SMOTE**) [3]

- Generates synthetic samples for the minority class by interpolating between minority class samples.
- New samples are created by selecting one of the $k$ nearest neighbors and interpolating.
- Number of synthetic samples generated is proportional to the minority-to-majority class ratio.

# Methodology

## Samplers — ADASYN

### Adaptive Synthetic Sampling (**ADASYN**) [4]

- Enhances SMOTE by focusing on samples near the decision boundary of the classifier.
- Uses an internal $k$-nearest neighbors ($k$-NN) classifier to identify samples near the boundary.
- Generates synthetic samples based on the density distribution of the minority class.

# Methodology

Correlation analysis

## Correlation analysis: EEG vs. Acoustic features

In addition, **Pearson correlation coefficients** are computed between the EEG and acoustic features, comparing the $17 \times 81 = 1377$ audio features with each of the 62 EEG channels across all imagined speech segments in the dataset. This provides an estimate of how well each EEG channel predicts the resulting audio.

| Channel | T7 | FT7 | TP7 | FT8 | P3 | CP5 | T8 | P5 | P7 | C4 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Pearson r** | 0.2397 | 0.2343 | 0.2297 | 0.2291 | 0.2284 | 0.2282 | 0.2281 | 0.2280 | 0.2277 | 0.2263 |
| **p-Value** | 0.0434 | 0.0467 | 0.0550 | 0.0521 | 0.0579 | 0.0573 | 0.0528 | 0.0568 | 0.0571 | 0.0492 |

Table: **Top 10 highest mean correlations** and their corresponding *p*-values between the acoustic and EEG features in each of the 62 channels over all the imagined speech segments in the dataset.

# Methodology

## Correlation analysis

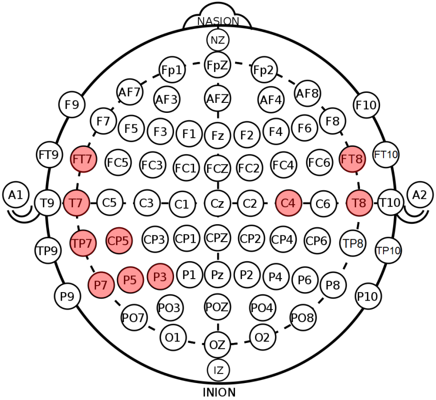


Figure: EEG electrode placement according to the International 10–20 system with the Modified Combinatorial Nomenclature (MCN). The red circles indicate the **top 10 EEG channels** with the highest absolute correlation coefficients between the **Acoustic** and **EEG** features.

# Results

## Performance metric analysis

Table: **Average accuracies (%)** across modalities and classes given the SVM-quad classifier from KaraOne methodology, and for select classifiers based on the proposed methodology.

| Method | Classifier | Modality | Task | | | | |
|---|---|---|---|---|---|---|---|
| | | | C/V | ± Nasal | ± Bilabial | ± /iy/ | ± /uw/ |
| KaraOne | SVM-quad | **EEG** | **18.08** | 63.50 | 56.64 | 59.60 | **79.16** |
| | | FAC | 62.54 | 48.10 | 63.73 | 40.25 | 20.68 |
| | | AUD | 81.05 | 40.48 | 39.98 | 37.63 | 18.33 |
| | | EEG+FAC | 72.17 | 48.41 | 63.73 | 56.03 | 19.60 |
| | | EEG+AUD | 61.13 | 62.72 | 39.99 | 49.15 | 83.75 |
| | | ALL | 75.72 | 51.87 | 63.73 | 46.01 | 20.20 |
| IFST | **RF + ROS** | **EEG** | **81.60** | 60.98 | 62.86 | 64.95 | **90.94** |
| | HC + NS / ROS | | 77.35 | 54.63 | 58.05 | 57.84 | **89.41** |
| | DT + ROS | | 69.55 | 55.61 | 55.96 | 61.25 | **82.65** |
| | SVM + ADASYN | | **75.47** | 44.74 | 54.91 | 56.03 | 40.14 |

# Results

Performance metric analysis

## Summary

- The Random Forest classifier with Random over sampler (**RF+ROS**) achieved average accuracies of $60\% - 90\%$, significantly improving over KaraOne SVM classifiers $18\% - 79\%$.
- The proposed methodology outperforms KaraOne SVM classifiers in most tasks, with RF+ROS showing the best performance among all models **except Task 2 ($\pm$ Nasal)**.
- Overall, improved performance is observed compared to KaraOne baseline models, despite some preprocessing steps being omitted.
- Variation exists among classifiers and samplers, reflecting their ability to handle uncertainties in EEG data.

# Results

## Performance metric analysis

### Summary

- Different samplers impact classifier performance significantly, indicating the necessity of balancing classes with samplers.
- Random Forest performs best due to its ability to handle data uncertainties by averaging multiple decision trees, reducing variance and overfitting.

# Results

Performance metric analysis

## Summary

- The Hanman classifier with No Sampler (**HC+NS**) and Random over sampler (**HC+ROS**) show *identical metrics* due to the classifier computing the least possibilitic uncertainty among all the pairs of norm errors among all the classes. Since *ROS doesn't generate new samples* but only duplicates the existing ones randomly, the classifier doesn't see any new samples and hence the metrics are identical.
- SMOTE and ADASYN samplers induce significant changes, highlighting the Hanman classifier's *sensitivity* to dataset alterations through sampling.

# Results

## Limitations

### Limitations

- The KaraOne feature selection method shows variability in the attributes of selected features for each task, greatly influenced by the choice of $N$.
- Introducing new data can alter the set of selected features, often discarding many without contributing additional information.
- In contrast, the proposed methodology maintains a fixed number of attributes per feature, independent of $N$, utilizing information from all features across windows and channels.
- Unlike KaraOne, which requires access to speaking segments in trials, the proposed method focuses solely on imagined speech segments.

# Results

## Limitations

### Limitations

- **Restricted Vocabulary Size:** EEG-based BCI systems for imagined speech classification are limited in recognizing a wide vocabulary due to subtle neural signals easily confounded by noise and overlapping mental activities.
- **Mental Repetition Challenges:** Participants often repeat the same word or phrase mentally for data generation, leading to mental fatigue and reduced signal quality, thereby impacting BCI system performance.
- **Individual Variability:** EEG signals vary significantly across individuals, necessitating personalized models that may be resource-intensive and challenging to generalize.

# Results

## Applications

### Application scenarios

Potential applications include:

- Brain-computer interfaces for communication in military settings.
- Assistive technologies for individuals with speech impairments due to neuro-biological disorders such as Alzheimer's disease, Parkinson's disease, and Amyotrophic lateral sclerosis (ALS), where *physical movement is impaired* but *cognitive function remains intact* [1].

# Conclusion

Summary

## Summary

- Explored Information set theory techniques to extract rich **spatio-temporal features** from EEG signals for imagined speech classification.
- Utilized features that significantly **reduce the dataset size** for training, while maintaining classification performance.
- Addressed the challenge of **effectively utilising all information** present in EEG signals without increasing computational complexity.

# Conclusion

Future scope

## Future scope of work

- Collecting **more data**
  - EMOTIV EPOC+ headset
- **Deep Learning** models
  - Evaluating the effectiveness of the IFS features
  - Possible with more data
  - Transfer Learning
- More **tasks**
  - **Inter-subject** variability
  - **Inter-session** variability
- **Multi-class** classification
  - Possible with more data
  - Each phoneme is a class

# References I


Mariana P. Branco, Elmar G. M. Pels, Ruben H. Sars, Erik J. Aarnoutse, Nick F. Ramsey, Mariska J. Vansteensel, and Femke Nijboer.

Brain-computer interfaces for communication: Preferences of individuals with locked-in syndrome.

*Neurorehabilitation and Neural Repair*, 35(3):267–279, 2021.

PMID: 33530868.

Leo Breiman.

Random forests.

*Machine Learning*, 45(1):5–32, Oct 2001.

Nitesh Chawla, Kevin Bowyer, Lawrence Hall, and W. Kegelmeyer.

Smote: Synthetic minority over-sampling technique.

*J. Artif. Intell. Res. (JAIR)*, 16:321–357, 06 2002.

# References II


Haibo He, Yang Bai, Edwardo A. Garcia, and Shutao Li.

Adasyn: Adaptive synthetic sampling approach for imbalanced learning.

In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 1322–1328, 2008.

Ashwin Kamble, Pradnya H. Ghare, Vinay Kumar, Ashwin Kothari, and Avinash G. Keskar.

Spectral analysis of eeg signals for automatic imagined speech recognition.

*IEEE Transactions on Instrumentation and Measurement*, 72:1–9, 2023.

J. Satheesh Kumar and P. Bhuvaneswari.

Analysis of electroencephalography (eeg) signals and its categorization–a study.

*Procedia Engineering*, 38:2525–2536, 2012.

INTERNATIONAL CONFERENCE ON MODELLING OPTIMIZATION AND COMPUTING.

# References III


Diego Lopez-Bernal, David Balderas, Pedro Ponce, and Arturo Molina.

A state-of-the-art review of eeg-based imagined speech decoding.

*Frontiers in Human Neuroscience*, 16, 2022.

Mamta and Madasu Hanmandlu.

Robust authentication using the unconstrained infrared face images.

*Expert Systems with Applications*, 41(14):6494–6511, October 2014.

Jeevan Medikonda, Saurabh Bhardwaj, and Hanmandlu Madasu.

An information set-based robust text-independent speaker authentication.

*Soft Computing*, 24(7):5271–5287, Apr 2020.

J. R. Quinlan.

Induction of decision trees.

*Machine Learning*, 1(1):81–106, March 1986.

# References IV


Jerry J. Shih, Dean J. Krusienski, and Jonathan R. Wolpaw.

Brain-computer interfaces in medicine.

*Mayo Clinic Proceedings*, 87(3):268–279, March 2012.

Scott Wellington and Jonathan Clayton.

Fourteen-channel EEG with Imagined Speech (FEIS) dataset, November 2019.

Shunan Zhao and Frank Rudzicz.

Classifying phonological categories in imagined and articulated speech.

In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 992–996, 2015.

Rümeysa İnce, Saliha Seda Adanır, and Fatma Sevmez.

The inventor of electroencephalography (eeg): Hans berger (1873–1941).

*Child's Nervous System*, 37(9):2723–2724, March 2020.