Predicting Human Brain Activity Using fMRI Data

Group no. 13

Punyam Singh
2110110842
Shiv Nadar University
ps232@snu.edu.in

Kartheek Kotha
2110110292
Shiv Nadar University
kk746@snu.edu.in

Bharathi Subramanian
2110110162
Shiv Nadar University
bn871@snu.edu.in

***Abstract*—This study explores the possibility of predicting human thoughts based on functional Magnetic Resonance Imaging (fMRI) data. Inspired by Tom Mitchell's work, we aim to reverse the process and predict words/categories a person is thinking about using fMRI data. The project involves data preprocessing, feature extraction, and model development. Results indicate promising insights into the relationship between brain activity and semantic content.**

I. Introduction

A. Background

Functional MRI (fMRI) has emerged as a powerful tool for investigating the intricacies of the human brain, allowing researchers to discern active regions associated with various cognitive functions such as speech, vision, and motor behaviour. The dynamic nature of the human brain in higher-order processes like language and reasoning has prompted the exploration of its activities through fMRI. Brain-computer interfaces (BCI) have already demonstrated the tangible link between brain activity and physical actions, enabling remarkable applications for individuals with disabilities. This project seeks to push the boundaries further by delving into the language and semantic domain. By leveraging fMRI data, the study aims to unravel the systematic relationship between brain activity and the intricate realm of human thought.

B. Problem Statement

At the core of this exploration lies a fundamental question: Can the rich information embedded in fMRI data be harnessed to predict the specific words or categories occupying an individual's thoughts? While existing research, exemplified by Tom Mitchell's work in 'Predicting Human Brain Activity Associated with the Meanings of Nouns' (Science, vol 320, 30 May 2008), has successfully modelled brain activity given the word, this project endeavours to reverse this process. The challenge is to predict the mental landscape—words or categories—from fMRI data, thereby opening new avenues for understanding the complex interplay between brain function and linguistic cognition.

C. Related Work

The foundation of this project rests upon prior research that has paved the way for understanding the intricate interplay between language semantics and neural activity, as exemplified by the groundbreaking work of Tom Mitchell and his colleagues. Their research primarily aimed at predicting human brain activity associated with the meanings of nouns, providing a computational model that has significantly shaped the landscape of neuro-semantic investigations.

As outlined in their research article, Mitchell's model presents a novel approach to predicting fMRI neural activation for arbitrary concrete nouns. The key methodology involves the utilization of twenty-five intermediate semantic features derived from sensory-motor verbs. This model is designed to learn the intricate mapping between these features and the observed fMRI images, utilizing a training set consisting of sixty words and their corresponding brain activation patterns.

The robustness of the model is underscored by its high accuracy in matching words to their fMRI images, highlighting efficacy within and across semantic categories. Notably, the model's predictions are not confined to the training set, as its applicability extends to previously unseen fMRI data for unfamiliar words, reinforcing its generalizability and potential for broader applications.

The outcomes of Mitchell's research offer profound implications for understanding the neural underpinnings of word meaning. By establishing a direct and predictive relationship between word co-occurrence statistics and neural representations, the model contributes significantly to unravelling the complexities of how the brain encodes concrete nouns. The model's ability to reveal learned neural signatures for each semantic feature further enriches our comprehension, shedding light on the neural components associated with sensory motor.

In the broader context of related work, Mitchell's contributions serve as a cornerstone for subsequent studies delving into the intersection of language semantics and neural activity.

II. METHODOLOGY

A. Data Collection

The fMRI data utilized in this study were obtained from the research article by Tom Mitchell et al., "Predicting Human Brain Activity Associated with the Meanings of Nouns" (Science, vol 320, 30 May 2008). The data are organized into separate files for each of the nine individuals, with each file stored in the MATLAB (.mat) format. Upon conversion to Pandas dataframes, the structure reveals a dictionary with keys (meta, info, data), where each key provides crucial information:

- `meta.study`: Indicates the name of the fMRI study.
- `meta.subject`: Serves as the identifier for the human subject.
- `meta.ntrials`: Represents the number of trials in the dataset.
- `meta.nvoxels`: Specifies the number of voxels (3D pixels) in each image.
- `meta.dimx,meta.dimy,meta.dimz`: Define the maximum coordinates in the x, y, and z dimensions, respectively.
- `meta.colToCoord(v)`: Provides the geometric coordinate (x, y, z) of the voxel corresponding to column v in the data.
- `meta.coordToCol(x, y, z)`: Determines the column index of the voxel with coordinate (x, y, z).
- `info.cond,info.cond_number`: Describes the condition (category of the word) and its numeric index during a trial.
- `info.word,info.word_number, info.epoch`: Specifies the presented word, its numeric index, and the number of times it has been presented during a trial.
- `data`: Contains the raw observed fMRI data, organized as a sequence of images collected over time.

Additionally, the study incorporates conversion functions from voxel coordinates to MNI coordinates and, subsequently, to AAL labels. These functions provide essential mappings for linking spatial information in the fMRI data to standardized brain coordinates and anatomical regions.

We also extracted the following data and used it in our approach to reverse engineer the process mentioned in Tom Mitchell's paper.

- `Intermediate Semantic Features`: These features, extracted from the supplementary website for Tom Mitchell's paper, provide the co-occurrences of each word according to a large text corpus. The dataset includes vectors of size twenty-five for all sixty words.
- `Word Embeddings using GloVe-Twitter-25`:
  Utilize newer technologies along our reverse engineering approach, we leveraged GloVe-Twitter-25, a pre-trained word2vec model, and extracted word embeddings of size twenty-five for each word. This step enhanced the dataset with richer semantic representations derived from the broader linguistic context captured by the word2vec model.

B. Data Preprocessing

Data preprocessing is a critical step in preparing the fMRI data for subsequent analyses, involving the conversion of voxel coordinates to MNI coordinates and the assignment of anatomical labels to each brain region. The process is facilitated by two key functions: `get_mni_coordinates` and `get_aal_labels`.

- Conversion to MNI Coordinates:

  The first step involves transforming voxel coordinates to MNI (Montreal Neurological Institute) coordinates using the `get_mni_coordinates` function. This function employs a transformation matrix, denoted as `transformToXYZmm`, to convert voxel coordinates (in voxel space) to MNI coordinates (in millimetres). The transformation matrix ensures a precise mapping of spatial information from the acquired fMRI data to the standardized MNI space.

  The transformation process includes appending a column of ones to the voxel coordinates, creating homogeneous coordinates. The transformation is then performed by matrix multiplication, yielding MNI coordinates for each voxel in the dataset.

```
transformToXYZmm
[[-3.125,  0,      0,   81.250  ],
 [0,       3.125,  0,  -115.625 ],
 [0,       0,      6,   -54.000 ]],
```

- AAL Label Assignment

  The second crucial step involves assigning Automated Anatomical Labelling (AAL) labels to each MNI point using the `get_aal_labels` function. This function utilizes an AAL atlas, comprising voxel inversion matrices (`vinv`), voxel coordinates (`Y`), labels (`label`), and unique identifiers (`id`). The AAL atlas provides a reference for mapping MNI points to specific brain regions. For each MNI point in the dataset, the process involves rounding the coordinates, applying the voxel inversion matrix, and extracting the corresponding AAL label from the atlas. The resulting list of AAL labels represents the anatomical classification for each voxel in the fMRI data. This data preprocessing pipeline seamlessly integrates the `get_mni_coordinates` and `get_aal_labels` functions, incorporating the transformation matrix to convert voxel coordinates to MNI coordinates accurately.

The resulting dataset contains both spatial and anatomical information, laying the foundation for robust analyses of the intricate relationship between brain activity and linguistic cognition. Post-preprocessing, the dataset comprises 117 unique labels, including the category "Not labelled," providing a nuanced characterization of the neural landscape under investigation.

The intermediate semantic features data extracted from the website required data cleaning such as string manipulations, filling empty cells, etc to be converted into a useful dataframe.

C. Feature Extraction

- Data Reduction based on Labels:

  The process begins with a meticulous grouping of data based on Automated Anatomical Labelling (AAL) assignments to each voxel. By calculating the mean of the sum for each experiment within specific labels, a structured 360 x 117 data frame emerges, where rows represent experiments, and columns signify distinct AAL labels. This initial aggregation sets the stage for subsequent analyses, providing a comprehensive overview of how different brain regions respond to experimental stimuli.

  Activation Values in fMRI:

  Functional Magnetic Resonance Imaging (fMRI) captures neural activity through voxel-wise intensity values, reflecting changes in blood flow and oxygenation levels. These values constitute the raw data, forming the basis for understanding brain responses. Summing these values across experiments unveils patterns of overall activation, offering a nuanced perspective on the neural dynamics associated with different AAL labels.

  Identification of High-Activation Labels:

  From the aggregated dataset, labels that significantly contribute to high activation values are identified. The summation of values across 360 experiments reveals labels demonstrating pronounced activation patterns. This step serves as a critical step to hoping in Following the identification of high-activation labels, the focus narrows down to extracting the most influential contributors. These labels, indicative of heightened neural responses, form a refined subset for more targeted analysis on specific brain regions that robustly engage with the experimental stimuli. The selection process ensures that subsequent investigations delve into brain regions most crucial for decoding the intricate relationship between linguistic cognition and neural activity.

  Voxel Data Filtering:

  To further refine the dataset, emphasis is placed on retaining only the voxel data corresponding to the selected contributing labels. This strategic filtering significantly reduces the dimensionality of the dataset, transforming it from a broader 320 x 21,764 to a more focused 320 x 5,611. The retained voxel data are specifically localized within brain regions identified as paramount for the cognitive processes under investigation. This feature extraction process ensures that the dataset is to capture the most informative elements for subsequent modelling and analysis.

- Semantic Feature Vector Mapping:

  In the data analysis phase, the semantic feature vectors corresponding to each word in the clinical trial context were meticulously mapped to their respective fMRI datasets. This mapping served as a crucial step to intertwine linguistic information with neural responses, fostering a more integrated understanding of cognitive processes. Enhance data quality, an innovative approach was adopted by computing the average semantic feature value for each word category. This aggregation process aimed to reduce noise and enhance the robustness of the dataset. By mapping these averaged semantic feature values to diverse types of fMRI datasets, we ensured a more nuanced representation of the cognitive nuances associated with each clinical trial.

- Word2Vec Model for Word Vector Extraction:

  Leveraging the Word2Vec model, word vectors with an array size of twenty-five were extracted for each word in the clinical trial descriptions. Word2Vec is a popular natural language processing (NLP) technique used to represent words as continuous vectors in a high-dimensional space. Word2Vec captures semantic relationships between words by training a neural network on large text corpora. The model we used was pretrained on Twitter's data consisting of 27 billion tokens. This approach provided a comprehensive numerical representation of the semantic content associated with each word. Recognizing the potential for mathematical operations on word vector data, a subsequent step involved extracting the average values of each data field for all word vectors sharing the same conditional number. This strategy not only streamlined the dataset but also facilitated more meaningful comparisons and analyses, laying the groundwork for a nuanced exploration of semantic and neural associations.

D. Model Development

In the quest to predict categorical outcomes across diverse datasets, a versatile model development pipeline was implemented.

1. MinMax Scaling: Adapted to the unique characteristics of the data by employing Min-Max scaling, ensuring consistent and effective preprocessing across different datasets.
2. KFold Cross-Validation: Ensured robust evaluation with a K-Fold cross-validation strategy (K=5), promoting reliable model performance assessment.
3. Model Evaluation:
   Diverse classifiers were systematically evaluated.
   i. LDA: Linear Discriminant Analysis captures linear relationships in the data, making it suitable for discerning patterns in the voxel intensity values.
   ii. Multi-Class LR: Logistic Regression, serving as a baseline, offers simplicity and interpretability, making it a benchmark for initial model assessment.
   iii. KNN: K-Nearest Neighbours explores local patterns in the data, potentially capturing localized relationships between voxel values and category predictions.
   iv. SGDC: Stochastic Gradient Descent Classifier adapts well to large datasets, offering efficiency in learning from the voxel intensity features.
   v. RFC: Random Forest Classifier excels in capturing complex relationships within the data, potentially uncovering intricate patterns in fMRI voxel values.
   vi. NB: Gaussian Naive Bayes Classifier assumes independence between features, providing simplicity and efficiency, which might be beneficial in the context of voxel-based predictions.
   vii. SVM: Support Vector Machines aim to find a hyperplane that best separates distinct categories, making it powerful in discerning nuanced relationships in the voxel data.
4. Results Visualization: Utilized the plot results function to generate visual summaries, offering a quick and informative comparison of model accuracies through horizontal bar plots.

This streamlined pipeline, emphasizing cross-validation, scaling, model evaluation, and visualization, stands as a flexible framework adaptable to diverse datasets.

In our parallel approach to reverse engineer the process mentioned in Tom Mitchell 's paper, we developed a model which accommodated the semantic features and word embeddings according to the following steps:

1. Normalization: Normalized the semantic features' data to have uniform sum of features for each word to ensure consistency and standardization among datasets.
2. Cross-Validation: We incorporated a combination of k-fold and leave-p-out (lpout) cross-validation techniques to ensure a robust and reliable evaluation of our linear regression model. K-Fold Cross-Validation was done using the same parameters as the previous approach.
   Additionally, we employed leave-p-out cross-validation, a variation of k-fold where p samples are withheld for validation in each iteration. This approach offers a more granular assessment, particularly beneficial when dealing with limited data. This helped us mitigate the risk of overfitting and provide a more accurate estimate of the model's generalization performance.
3. Models Used:
   Two distinct set of response variables were used:
   i. Linear Regression: We used a Multi Output Linear Regression Model which used different datasets as predictor variables such as voxel data of different people, fMRI data based on AAL labels, combined data of all nine people, etc. The response variable was the average value of semantic features grouped by category, harnessing the unique relationship between words and semantic features, and the representability of words as intermediate semantic features.
   ii. Linear Regression: We used a Multi Output Linear Regression Model which used different datasets as predictor variables such as voxel data of different people, fMRI data based on AAL labels, combined data of all nine people, etc. The response variable was the average value of the word vectors data grouped by category. The predictor variable in this model has captures the relationship between words better than semantic features as the word2vec model is better at representing words and allows mathematical operations to the word vectors.
4. Model Evaluation:
   i. R^2 Score: We employed the R2 score in assessing our past linear regression model to quantify the proportion of variance in the predicted values explained by the model.
   ii. Mean Squared Error: Mean squared error was employed in the linear regression model to quantify the average squared difference between predicted and actual values.

III. Experiments And Results

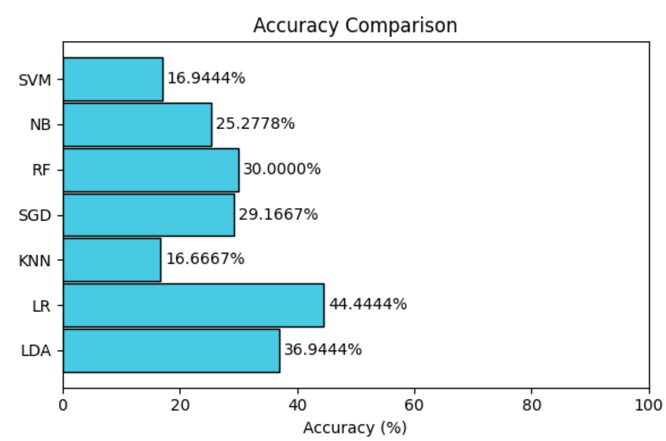
In the pursuit of decoding the neural underpinnings of linguistic cognition, a series of experiments were conducted, exploring various facets of feature representation and model performance. The results, derived from a comprehensive analysis pipeline, shed light on the intricate relationship between brain activity and semantic processing.

First Person Analysis:

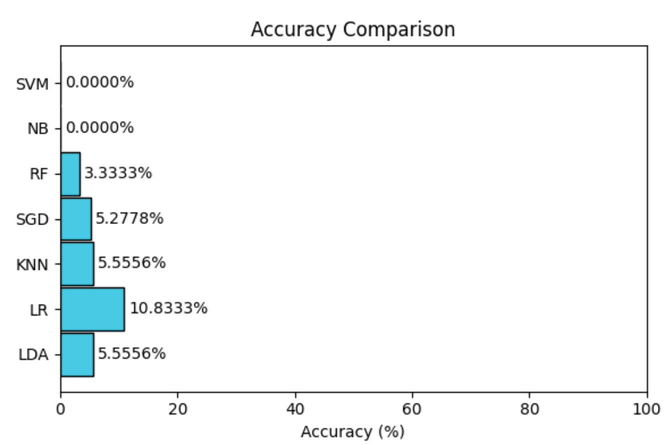
1. All Voxels:

For the first person, the initial exploration considered all voxels in the dataset, this encompassed a meticulous examination of categories and words, employing n-fold cross-validation across diverse models. Additionally, the integration of word vectors and word semantic provided a nuanced perspective on word representations. The detailed analysis aimed to uncover patterns of neural activation associated with specific semantic content.

Categories: We have 12 distinct categories for 360 samples



Words: We have 60 distinct words for 360 samples, where each category has 5 words



When average values of word vectors, grouped by categories were used as response variables, we got the following evaluation metric results:

Mean Squared Error: 0.11243

R-squared: 0.24409

2. Grouped Data:

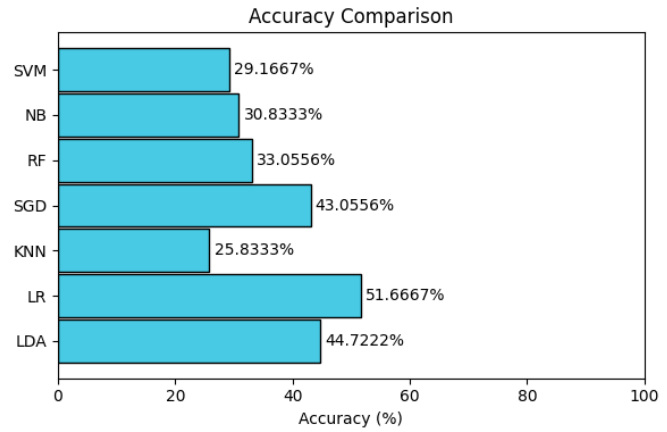
In the context of grouped data, the investigation delves into the intricate interplay between neural activation patterns and specific anatomical regions. The meticulous grouping of data based on Automated Anatomical Labelling (AAL) assignments allows for a nuanced exploration of how different brain regions respond to experimental stimuli. By calculating the mean of the sum for each experiment within these specified labels, a structured 360 x 117 dataframe is crafted. Each row in this dataset encapsulates an experiment, while the columns represent distinct AAL labels. This initial aggregation lays a robust foundation for subsequent analyses, providing a comprehensive overview of the neural dynamics associated with different anatomical regions.

The identification of high-activation labels becomes a pivotal aspect of this exploration. From the aggregated dataset, labels that significantly contribute to elevated activation values are discerned. This process involves the summation of values across 360 experiments, revealing labels characterized by pronounced and consistent activation patterns. This identification sets the stage for a more refined subset of labels, representing the most influential contributors to neural responses. This subset becomes pivotal for a more targeted analysis, focusing specifically on brain regions crucial for decoding the intricate relationship between linguistic cognition and neural activity. The strategic voxel data filtering that follows ensures that the dataset is honed to capture the most informative elements, reducing dimensionality from 320 x 21,764 to a more focused 320 x 5,611, and refining the dataset for subsequent modelling and in-depth analysis.

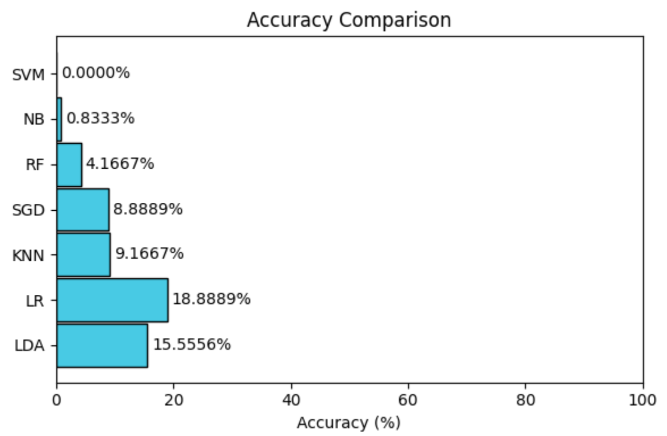
3. Selected Voxels:

In a more targeted approach, the analysis narrowed down to selected voxels based on high-activation labels. This refined subset aimed to capture the most relevant neural signatures for semantic processing. The examination included n-fold cross-validation for categories and words, encompassing various models, including. Additionally, the integration of word vectors and word semantic provided a nuanced perspective on word representations. The detailed analysis aimed to uncover patterns of neural activation associated with specific semantic content.

Categories: We have 12 distinct categories for 360 samples

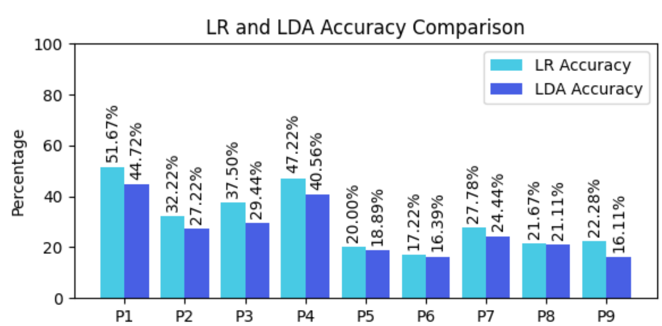


Words: We have 60 distinct words for 360 samples, where each category has 5 words



All Persons Analysis - Discrete Persons:

Extending the investigation to include data from all individuals, the analysis focused on discrete features, employing a min-max scaler for normalization. This broader exploration aimed to uncover patterns that transcended individual variability, providing a more generalized perspective on the relationship between brain activity and semantic processing.

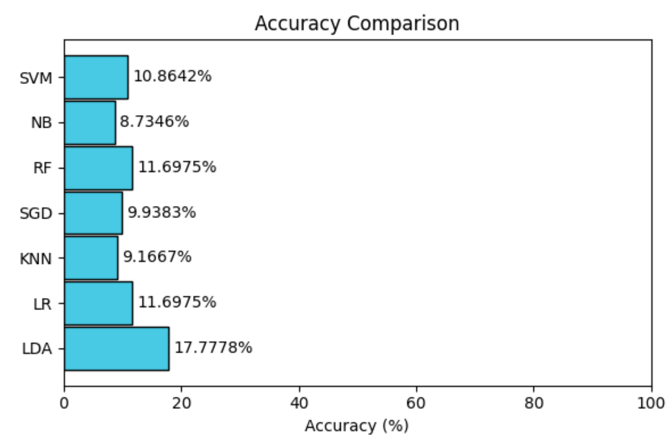


Combined Persons Analysis - Problems and Observations:

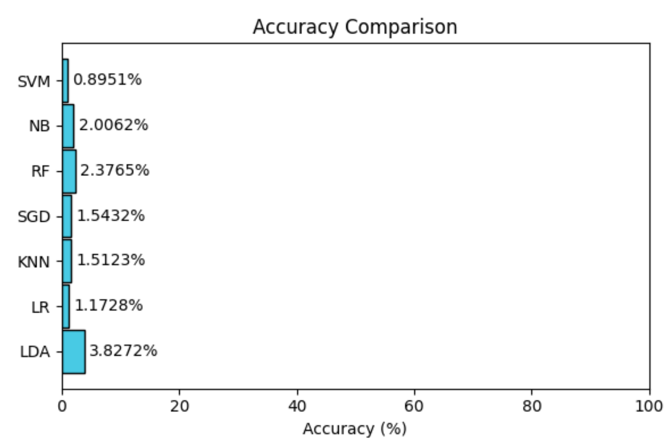
The integration of data from multiple individuals brought about both challenges and valuable observations. While combining data allowed for a broader understanding, it also presented challenges related to individual variability. Addressing these challenges required thoughtful considerations to ensure the robustness and generalizability of the results.

Challenges and Observations:
1. Consistent Experimental Setup: All subjects underwent identical experimental procedures, ensuring uniformity in the conducted experiments.
2. Varying Voxel Counts: The subjects exhibit variations in the number of voxels, posing a challenge in consolidating their data into a singular dataframe due to the differing dimensions.
3. Subject-Specific Label Diversity: Each subject possesses a distinct set of unique labels, making the utilization of ungrouped data impractical. To make it uniform we have added the missing labels to each dataset and initialized it to zero (meaning no image intensity).

Categories: We have 12 distinct categories for 360 samples



Words: We have 60 distinct words for 360 samples, where each category has 5 words.



Word Vectors: We extracted the word vector expressing the semantic relationship between words using Word2Vec model.

Results obtained using above data:

Mean Squared Error: 0.15333

R-squared: 0.03421

In summary, the experiments and results presented herein provide a comprehensive exploration of the neural correlates of linguistic cognition. The insights garnered from individual and combined analyses offer a nuanced understanding of how brain activity reflects and contributes to the intricate processes underlying language comprehension.

V. Conclusion

In conclusion, our endeavour to predict words based on fMRI data has yielded multifaceted insights into the intricate relationship between neural activity and linguistic representations. By leveraging AAL labelled data, we navigated the complex terrain of brain regions, unravelling patterns indicative of word prediction. The strategic selection of labels based on activation values enhanced the specificity of our predictions, shedding light on the nuanced interplay between specific brain regions and semantic content. The integration of semantic features further enriched our understanding, bridging the semantic-neural gap and providing a comprehensive framework for cognitive exploration. The incorporation of word embeddings from the Word2Vec model allowed us to capture intricate semantic nuances, enabling a more refined analysis of the language-cognition interface. However, our journey is not without its limitations; variations in voxel data, insufficient word-related data, and other inherent inadequacies have constrained the attainment of exceptionally high accuracies. Acknowledging these challenges, we have heeded our professor's recommendation to incorporate word vectors, a decision that proved pivotal in enhancing the interpretability and depth of our findings.

VI. Future Work

In the pursuit of advancing our predictive models for fMRI data analysis, several avenues warrant exploration. Firstly, a comprehensive extension of neural network applications to cover the entire dataset is essential. This entails a thorough investigation of various neural network architectures and hyperparameter configurations to ascertain their effectiveness across diverse subjects.

Another promising avenue involves the transformation of vector data into image vectors, followed by the application of Convolutional Neural Networks (CNNs). This innovative approach aims to exploit spatial relationships inherent in the transformed data, potentially unlocking more intricate patterns that could enhance predictive accuracy.

Furthermore, there is a need for a refined strategy in leveraging word vectors across all subjects' data. This involves an in-depth exploration of advanced methodologies to better utilize word representations, seeking to optimize their contribution to model performance. Consolidating efforts in these directions will contribute to a more comprehensive and effective analysis of fMRI data.

VII. References

1. https://www.science.org/doi/10.1126/science.1152876
2. https://www.cs.cmu.edu/~tom/science2008/index.html
3. https://www.cs.cmu.edu/~tom/science2008/featureSignaturesP1.html
4. https://www.sciencedirect.com/science/article/pii/S1053811919307803
5. https://radimrehurek.com/gensim/models/word2vec.html

VIII. Acknowledgments

The authors express their gratitude to Professor Harish Karnick for guidance and support throughout the project. Special thanks to Tom Mitchell for providing access to the fMRI data through his esteemed research paper on Predicting Human Brain Activity Associated with the Meanings of Nouns.

IX. Supporting Resources

https://github.com/punyamsingh/PredictingHumanBrainActivity