Visual Object Classification from fMRI Data

by

Syed Mishar Newaz
18101210
Taslim Ahmed Taseeb
18101443
Abdullah Nurul Haque
18101694

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2022

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mishr

Syed Mishar Newaz
18101210

Taseeb

Taslim Ahmed Taseeb
18101443

Nhaque

Abdullah Nurul Haque
18101694

Approval

The thesis/project titled “Visual object classification from fMRI data” submitted by

1. Syed Mishar Newaz (18101210)
2. Taslim Ahmed Taseeb (18101443)
3. Abdullah Nurul Haque (18101694)

Of Fall ’21 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 16, 2022.

Examining Committee:

Supervisor:
(Member)

Zavid Parvez

Digitally signed by Zavid Parvez
DN: cn=Zavid Parvez, o=Brac University, ou=CSE, email=zavid.parvez@bracu.ac.bd, c=US
Date: 2022.01.20 12:35:34 +11'00'

Mohammad Zavid Parvez, PhD
Assistant Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Computing devices were once limited in just calculating arithmetic. Whereas, in modern computing, complex task like object classification or recognition has become so popular that even our smart devices cannot be thought without having a voice, character and face recognition features. Although it has been a long time since the idea of object recognition first came into the scene, there has been limited amount of work done in categorising objects from human fMRI data. As a result, part of human cognitive study has been neglected which possesses a large potential to be discovered and used. In brief, when a human perceives an object through vision or imagination, certain regions of brain generate specific patterns of electric signals. Using $fMRI$ brain data, we can potentially use those signals to interpret whatever a person is perceiving. We have tried to recreate some of the few works done previously in a limited test environment. In this paper, we try to explore an approach where a random perceived object gets split into a bunch of features it possesses. Using those extracted features, we will be able to classify the object from our previously trained deep learning model. Finally, our experiment will show a robust approach to explore and study human cognition using computers.

Keywords: Functional MRI; Machine Learning; Visual Features; Convolutional Neural Network; Deep Learning

Table of Contents

List of Figures

List of Tables

Nomenclature

This list mentions some symbols and acronyms which has been used in this paper

CNN	Convolutional Neural Network
$fMRI$	Functional Magnetic Resonance Imaging
$T2B$	Top To Bottom
FFA	Fusiform Face Area
GIST	Generalized Search Tree
HMAX	Hierarchical Model and X Pooling
HVC	Higher Visual Cortex(LOC, PPA, FFA)
LOC	Lateral Occipital Complex
LVC	Lower Visual Cortex(V1, V2, V3)
MVPA	Multi-voxel pattern analysis
PPA	Parahippocampal Place Area
ROI	Region of Interest
SIFT	Scale-invariant Feature Transform
VC	Visual Cortex

Chapter 1

Introduction

1.1 Introduction

One of the fundamental approaches of predicting human cognitive activities is the analysis complex patterns from functional (MRI), which stands for Magnetic Resonance Imaging. These includes whatever a person perceives while watching, thinking, memorizing as well as dreaming during his sleep. However, simple linear classification based approaches yield results which are not effective in the real world application. This is due to the limited capabilities of the linear classifiers. On the other hand, modern techniques involve encoder/decoder based model which generates possible brain signal database from image. Then, the signals are compared with the actual $fMRI$ data to predict the brain activity. While this approach is great for image identification, it lacks the ability to provide rigid details whether the object a person is watching with eyes, just imagining in the brain or just dreaming. As the amount of objects a human observes in day-to-day life are limitless, we need a better approach to categorize those. This paper aims to establish a novel approach for categorising generic objects by decoding brain activity of subjects who either watched or just imagined an object in their head.

1.2 Problem Statement

Human race is different from machines due to their diverse ability to perceive, think, dream, infer and classify objects. This intelligence has given birth to innovations. As a result, naturally they can extract relevant features to classify an arbitrary object and continue improving the capability to classify new objects. However, machines are different. So-called Machine vision is just a set of sampled pixels with numeric data. Also, machines do not have an inherent ability to think or classify objects. Due to that, we have to hard-code a machine program in order to do something reliably.

However, if we think for a bit, we can realize a problem with this approach. With the passing of time, the number of distinguishable objects in this universe are increasing at an exponential rate. If we stick with this approach, we have to hard-code a program to classify just one specific object. Obviously, that would require an infinite number of programs to be written and tested at every moment in order to classify an acceptable amount of the real-world objects reliably. This is the reason we needed a more robust yet generalized approach.

Observing these obstacles, we decided to build a system which will take $fMRI$ brain signals, then process and convert those signals into a real world object features. Using a trained model, those features will be used to categorise a real-world object. This approach will help people with speaking inabilities to express their statement, recall forgotten things and improve their lifestyle.

1.3 Objective

- **Learn about fMRI**
- **Explore real-world fMRI data**
- **Correlate fMRI signals with real-world objects**
- **Make an object classifier**
- **Understand human cognition better**

1.4 General Overview

Initially, we execute some pre-processing tasks on the $(fMRI)$ data inside the dataset. In this stage, we demonstrate that not every ROI of brain is relevant for us. Instead, we can narrow it down to specific signals obtained from certain parts of the brain. As we are mostly interested in visual features, our focus will be limited to regions related to visual regions of brain, such as Visual Cortex, Parahippocampal Place and Occipital Complex. This gives us some additional benefits as well. As $fMRI$ is extremely noisy data due to uncountable number of brain activities, narrowing the region gives us less signal to noise ratio. Additionally, we show that stimulus-trained decoders can be used to decode visual features of imagined objects, providing evidence for the progressive recruitment of hierarchical neural representations in a $T2B$ manner.

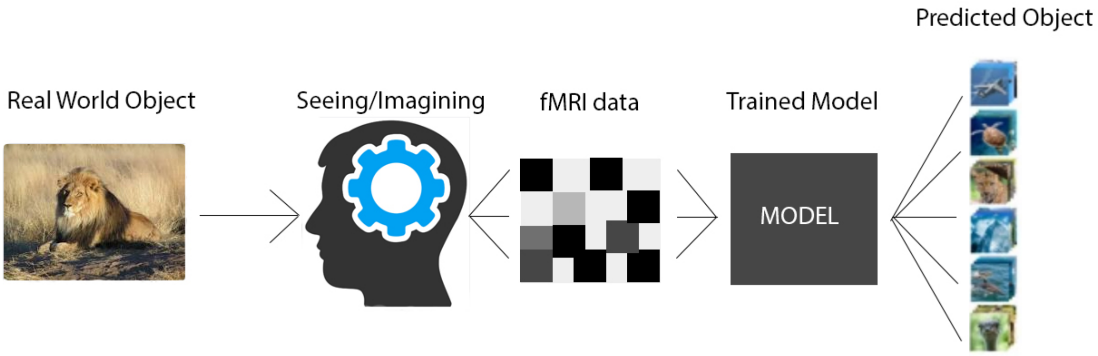

Figure 1.1: $fMRI$ data to object classifier overview

Last but not least, we evaluate our model by testing it and determine whether it is useful at all recognising arbitrary objects from brain signals. We try to maximise the accuracy of this model as much as possible. Although, we are in a challenging situation due to lack of computational power to train and evaluate multiple times in order to gain optimal efficiency. Therefore, we have to compromise in some areas in order to finish this study in the given amount of time. In future, these limitations can be overcome by improving those compromised parts of this study.

1.5 Thesis Orientation

This segment gives an overall overview of what has been discussed in each chapter of this thesis paper. After giving the overview of this study and what we intend to do and plan to accomplish in this chapter, the remainder of the paper is assembled in the consequential manner

- Chapter 2 reviews the relevant papers
- Chapter 3 presents the work plan and lab environments
- Chapter 4 mentions the methodology of this research
- Chapter 5 describes the implementation details and results
- Chapter 9 concludes this thesis with future plans

Chapter 2

Literature Review

While doing our research on image decoding and fMRI analysis, we came across some articles presenting interesting phenomena. These articles reflect what have been done previously.

Horikawa and Kamitani(2017)[7] had the objective of decoding arbitrary object categories. It was done from human brain activity and by using fMRI. So, four different computational models and 13 visual feature types/layers were used to build visual features from images. The models used were: CNN (8 layers), HMAX (3 layers), 1 layer of GIST, and SIFT+BoF each. The authors performed two fMRI experiments: experiment of image representation and experiment of imagery. In the image presentation experiment, a training image session was conducted. Every 1200 images from 150 categories were once presented there. In test image session, 50 images were presented. They were chosen from 50 object categories and were presented 35 times. At the time of imagery experiments, 1 of the 50 image categories was imagined by the subjects. fMRI signals were measured during that time. The target objects for analysis were: a) the models and b) ROIs. The authors trained a unit vector for predicting feature vector values. Around 1000 decoders were used per layer. These decoders predicted a feature vector when the subject was seeing or imaging an object which was left out during the time of training. Finally, for the identification of the seen/imagined object, the authors calculated the resemblance between vectors that were predicted and the feature vectors. As for accuracy, CNN1-8 showed the most (more than 85% for seen object identification and close to 70% for identifying imagined objects).

M. Johnson and M. Johnson(2014)[4] had the aim to discover what was performed in cortex areas related to scene selection and verify if they created any item-specific details at the time of mental imagery. They also wanted to find out the level at which information was represented for re-evaluating item-specific task motifs noticed during visual perception. Then, MVPA was used with the help of support vector machine(SVM) classifier. This was done to detect if during perception and/or image those regions could properly encode the information about the identity of specific scene items. After the experiments were completed, they found several scene-selective regions which were OPA, PPA, RSC, and PCR/IPS. Item-specific data was represented in these regions.

Miyawaki et al.(2009)[1] made an effort to reconstruct the visual images using multivoxel patterns of fMRI signals and multi-scale visual representation. To rebuild the presented image; at first, a decoder was used using multivoxel patterns to predict the stimulus state at each local element. Then finally the outputs of all local decoders were combined. This combination was used to represent several different complicated images. The subjects observed a sequence of visual images. These images consisted of binary contrast patches on a 10x10 grid. FMRI signals were measured during that time. Two image sessions were conducted. In "random image session", 440 different images, and in "figure image session", five alphabet letters and five geometric shapes were shown 6/8 times. The objective for analysis was fMRI signals from V1 to V4. "Random image session" data was used in the "figure image session". This was done to rebuild the images. This paper also performed image identification analysis along with image reconstruction to measure the accuracy. The identification was more than 95% accurate.

Naselaris et al.(2009)[2] constructed a Bayesian framework. They wanted to create the spatial structure of natural images with proper reconstructions for brain reading. Here the authors used the image with the highest posterior probability of creating the measured response as the definition of reconstruction. For calculating the probability, two details were used: a) target image details and b) prior details that contained the structure as well as semantic content of native images. The authors used an additional semantic encoding model to create reconstructions. These reconstructions properly reflected the semantic content of target images for the improvement of rebuilt images. These images were collected from fMRI data. To calculate semantic accuracy the authors considered semantic categories at four different levels of specificity. From 2 categories that were vividly defined, showed an accuracy of 90% to 23 narrowly defined categories that have an accuracy of 40% by the hybrid method were used.

Naselaris et al.(2011)[3] demonstrated a voxel-wise modeling and decoding approach. Their aim was to establish that while visualizing multiplex images from memory, visual features that were low-level were encoded in generated activity. The authors believed that the encoding model approach separates specific elements with a variation that occurred due to visual features with low-level. Then that element is used in the identification of mental image connected with a measured pattern of activity[5].

Based on magnetic resonance imaging (MRI) of the brain and 3D convolution neural networks(3D-CNN), a paradigm was presented for distinguishing individuals with schizophrenia from healthy control participants(Han et al., 2015) [6]. Using 3D-CNN based, ten-fold cross-validated deep learning classification framework and features based on ICA, resting-state functional MRI data from 144 participants were divided into two groups of 72, ageing from 18 to 65 years where one was filed with schizophrenia patients and the remaining 72 were healthy people serving as controls which was acquired from the COBRE dataset. Siemens TIM 3.0-Tesla Scanner was used to scan the patients with multi-echo MPRAGE (MEMPR) sequence and rs-fMRI data were obtained using single-shot full k-space echo-planar imaging (EPI) with ramp sampling correction where the intercommissural line (AC-PC) was used as a reference. They found out that their classification accuracy was 98.09 ± 1.01%, the p-value was less than 0.001 and AUC (which stands for Area under the curve) was 0.9982 give or take 0.015. Furthermore, upon statistically analysing across several resting-state networks dissimilarity in functional linkage among the two groups, it was revealed that patients had a better relationship between the default mode network and different cerebellar or networks that were task-positive but had a noticeable gap between the visual and frontal networks. For the determination of schizophrenia, these ICA functional network maps functioned as largely deterministic 3D imaging characteristics. Finally, it can be said that in the future, owing to its high AUC, this study might be used as an additional instrument to aid doctors at the first evaluation of schizophrenia in the future with further cross-diagnosis validation and a publicly accessible data set.

Hosseini et al.(2019)[10] in their paper outlined, constructed and applied an automatic computing BCI system for restriction and prediction of Epileptogenicity. They used rs-fMRI which stands for resting state-functional magnetic resonance imaging as well as electroencephalography (EEG) to analyse functional connectivity. They created and implemented both nonintrusive and intrusive approaches for monitoring, assessment, and regulation of the epileptic brain by taking advantage of autonomic edge computing in epilepsy. To overcome existing obstacles, novel methods were proposed for processing multimodal rs-fMRI and EEG big data which was obtained independently for epileptogenic network definition and prediction via deep learning. They used an unsupervised feature extraction model for recognising both preictal and non-preictal time periods of data through convolutional deep learning where a nonlinear support vector machine(SVM) with a GRBF kernel classifier was used. Also an edge computing framework for live and prevalent computing of big data found in medical fields such as fMRI,iEEG and EEG as part of a DSS for surgical suitability has been introduced. They modelled the brain as a connected mesh of nodes and connectivity matrices are approximated from rs-fMRI and EEG data. Wavelet analysis was performed to extract various features from epileptic EEG signals. They found out that the proposed approaches gave an accuracy of 98%, sensitivity of 96%, specificity of 97% and better p-values. These are supported by experimental and simulated findings based on real-world patient data. Although concurrent EEG and fMRI data acquiring may be difficult but the authors proposed models to overcome this computing over independently obtained EEG and fMRI data.

Qureshi et al.(2019)[9] in their paper introduces a new framework that detects arousal levels by combining sophisticated multimedia features obtained from a collection of sample videos and the functional activity of people's brain recorded by functional magnetic resonance imaging (fMRI) concerning the people's reaction to videos. Initially, 183 videos from various categories and arousal levels are used to build a database and 93 out of 183 videos are arbitrarily chosen fMRI scan's training dataset.10 males and 10 females with ages varying from 21-30 years willingly watched the videos and labelled them according to arousal levels 1-5. are used to tag the video arousal. Afterwards many audio(zero cross, roll-off, tempo, pitch and MFCCs) and video features(aesthetics, shot length feature, general preferences, visual excitement satu- ration, colour heat, and motion features)all of which are linked together to construct a composite feature vector and fed to the multimodal

DBM algorithm and this is used for training classifier where libSVM with RBF kernel was used. Next, fMRI- derived feature extraction takes place which consists of fMRI data acquisition by scanning(multimodal DTI and fMRI scans) 3 healthy participants separately by MRI system (GE 3T Signa HDx) having a head coil with eight channels, brain ROI identification where 358 predicted and consistent ROIs are used to form eigenvector which rep- resents fMRI signal of Roi, and feature extraction where functional connectivity is computed among a pair of ROIs by Wavelet transform coherence(WTC). As fMRI scans could prove to be non-economical and lengthy, DBM model can represent both audio and video features combined in absence of fMRI scans. They found out after carrying out the experiment that their integrative method inevitably provided improved accuracy(10% to 15% increase) compared to previous works. In future, they will be improving accuracy by using a better algorithm and working with wider ranges of data sets.

Han et al.(2020)[8] in their paper challenged the traditional ability of convolutional neural networks (CNNs) that are exclusively feedforward and driven by goals to anticipate and interpret cortical responses to natural pictures or videos. It achieves such a feat using variational auto-encoder (VAE) which serves as a different deep neural network. 1,024 latent variables and encoders and decoders consisting of five hidden layers were included in the VAE. It was trained using the ImageNet ILSVRC2012 dataset with training samples exceeding 2x106 unlabeled images and for training purposes, the Adam optimizer was used that had a learning speed of 10-4. For this study, there were 3 female participants, ranging from 23-26 years old and viewed video content which was of 13.7 hours duration. The encoders and decoders were trained from two independent data sets for both the prediction of fMRI response as well as video reconstruction. fMRI data of various spatial and temporal resolution were logged and preprocessed with the minimal preprocessing pipeline. It was found out that VAE in contrast to CNN, performed poorly in higher-order optical regions but performed on par with CNN when it came to forecasting cortical responses due to videos in prior optical regions. Upon investigation, it was found out that the main reason for this was their different training aims instead of differing model architecture or amount of parameters. When compared to CNN, VAE had equal accuracy in predicting video- Induced cortical feedbacks in prior visual regions, but worse accuracy in higher-order visual regions. The discrepancy in carrying out encoding between CNN and VAE was creditable mostly to their differing learning aims, instead of differing model architectures or the varying amount of parameters in them. VAE provided an easier scheme when it came to interpreting the fMRI activity to remake the inputted motion picture, through an initial transformation of the fMRI data into latent variables for Variational autoencoder, and later turning them to the rebuilt motion picture frames using VAE's decoder. It was evident that this procedure was much better than other methods for its ability to rebuild both the colour and spatial composition of the visual input. From these results, it becomes clear that although VAE has its merits and demerits when it comes to describing feedbacks from the cortex and remaking a plethora of life-like optical scenes, it can be an exemplary unsupervised model for learning visual representation.

Chapter 3

Work Plan

3.1 Overview

In our model, our goal was to extract some features from real world images which we would use to classify objects later. We used CNN to achieve that. First, $fMRI$ activities needed to be recorded and observed while the subject is seeing or imagining an object. Then, we needed to extract visual features from the activities using CNN and built a feature vector to train the model. Later, when a subject observes an image of an object, the model will be used to predict all the possible names of that observed object.

3.2 Environment

In order to train the model, we have used our personal computers along with google colaboratory in order to accelerate our processing. We used 2 separate personal computers for distributing the training workload. This also helped us in reducing the total computation time. In total, the model training took around 60 hours with all these performance tweaks applied. Without multiprocessing, this would have taken more than 160 hours.

3.3 Third Party Libraries

Some popular libraries have been used to accelerate our code writing. Notable ones are ”numpy”, ”pandas”, ”h5py”, ”bdpy”, ”multiprocessing”, ”pickle”, ”sklearn” etc. ”Numpy” is mostly used for advanced mathematical operation conveniently. It provides interface to calculate matrix operation with preset precision. ”Pandas” is used to organize data in data frame and perform operation in a declarative fashion. ”Bdpy” and ”h5py” is used for storing fMRI time series data and manipulating it. Pickle is used for dumping and restoring machine learning models without the need of re-training. ”Sklearn” is a library we used to do some model training and classification related operations.

3.4 Workflow Chart

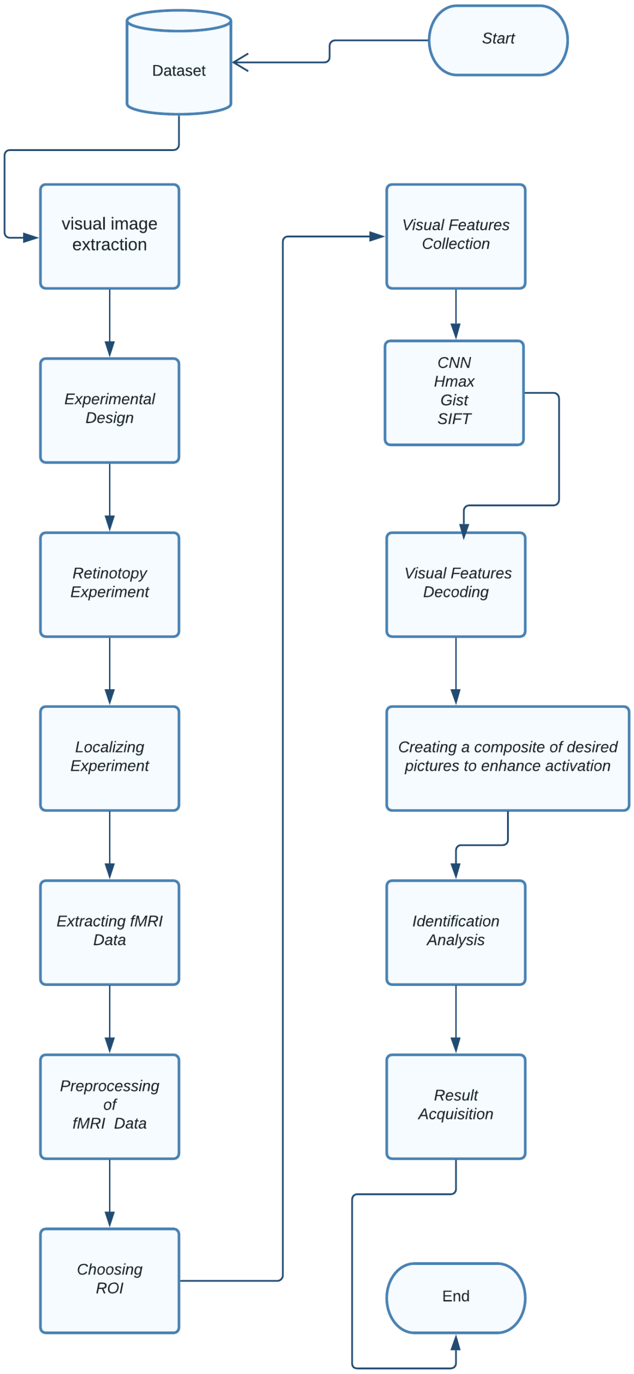

Figure 3.1: Workflow Chart

3.5 Gantt Chart

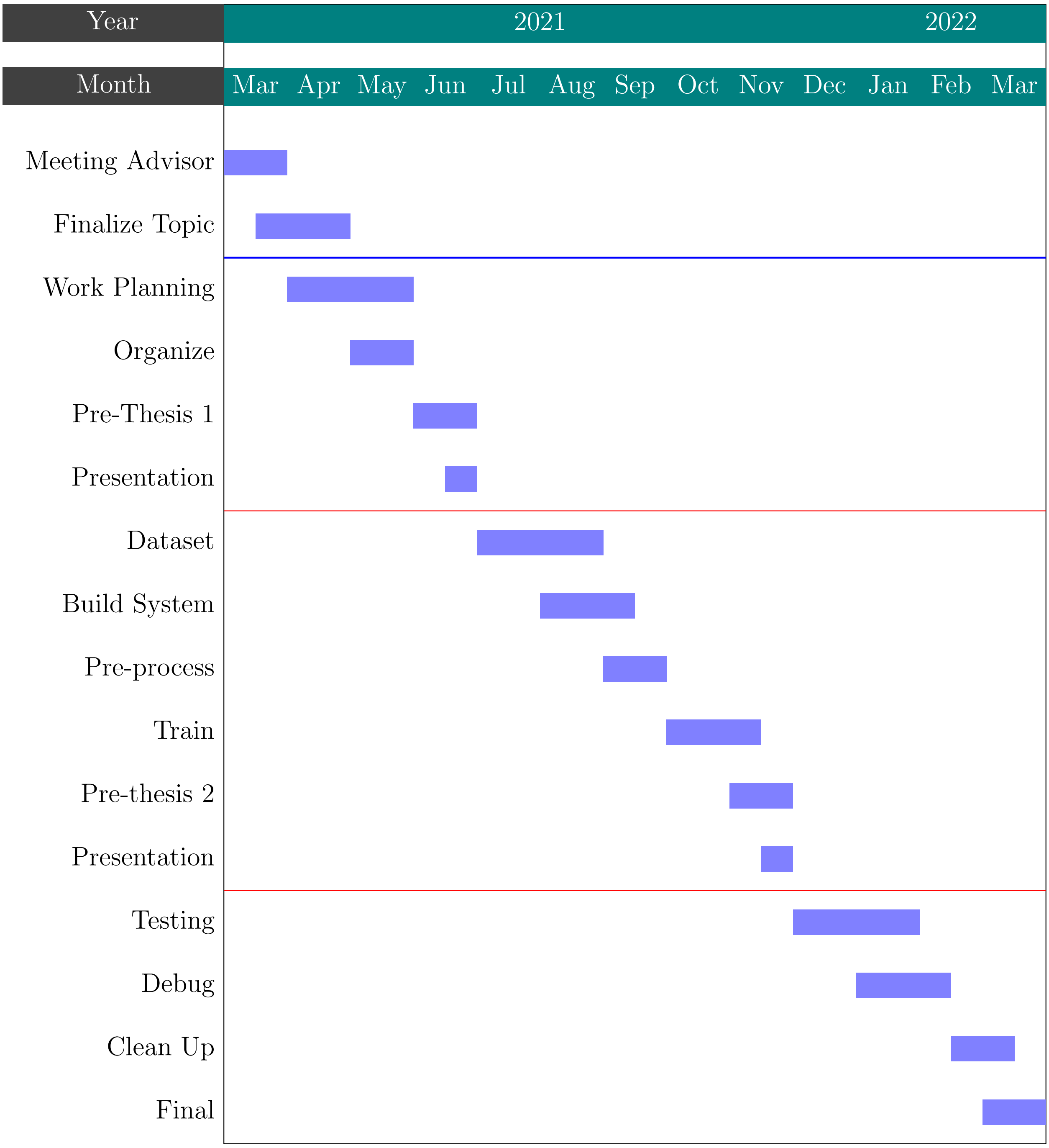

Figure 3.2: Workflow Gantt Chart

Chapter 4

Methodology

4.1 Overview

Computational models such as CNN, HMAX, GIST and SIFT were used to extract visual features from natural images. We multi-voxel fMRI data from the dataset. A set of features were predicted based on observed fMRI activity. Then it was utilised to predict the observed object. This identification task was done by comparing the current feature vectors with the feature vectors of other objects. The objects were stored in an annotated picture library which also included ones that were not used for decoder training.

4.2 Input Data

The input data along with some code snippets were collected from Kamitani LAB which consists of 5 subjects' fMRI recording data. The official dataset is in h5 format. The dataset contains corresponding fMRI data while the subject was exposed to visually witnessing an object or imagining it. In our experiment, we only focused on visual objects and their features. There were around 150 categories and 1200 classes, where 8 images were from the same category. It also contains a gigantic ImageNet (https://www.image-net.org) library of 15372 categories to later evaluate the model. However, as we extracted features only and used that to predict object, the categories are not limited to trained data. The brain data contained around 500 to 1000 voxels depending on the regions. Multiple *ROI*s are seperated using labels and datatype.

Array Name	Array Shape	Contents
Data	5 x 10 x 13 x 1000	Subject x ROI x Feature x Voxel
ImageID	150 x 1200	Category_Name x Class_Name

Table 4.1: Input Data

4.3 Pre-Processing

In order to read and interpret the data, we extracted the necessary data using h5py library. Then, we used bdpy data structure to seperate ROIs for particular stimulus along with its' corresponding data type. Also, we needed to filter the data as the noise:signal ratio was very big. The image feature unit was extremely big as well, which our computer was taking a long time to train. So, we used only 100 units of feature data for training. Then, we flattened the data and divided it into test:train with 60:40 split. This whole process was done using multiprocessing. At this point, the data is converted in a suitable form to run our analysis.

4.4 Feature Prediction

At this point, we ran iteration for all the of (subject x rois x features) combinations. For each combination, we started by normalizing the data. Then, we found the correlation coefficient. In order to find the coefficient, we used Sparse Linear Regression with 500 iterations in each combinations. However, for later trainings, we decreased the iterations to 5 as it was taking a lot of power from our CPU. After that, we initially used 8 layers of CNN with 1000 neurons in each, 3 Layers of $HMAX$, 1 layer of GIST and SIFT to train the model. However, later we had to reduce the number of layers to 3 due to lack of cumputational power. Anyhow, we distributed voxel sets throughout different layers. For example, LVC (lower visual areas) V1, V2, V3 were given to 2nd layer of CNN whereas arterior areas like LOC(Lateral Occipital Complex), PPA(Parahippocampal Place Area), FFA(Fusiform Face Area) were given to 8th Layer of CNN. Different algorithms performed better in terms of accuracy for different class decoding. The specific number of iterations and weights were adjusted after lots and lots of tweaks.

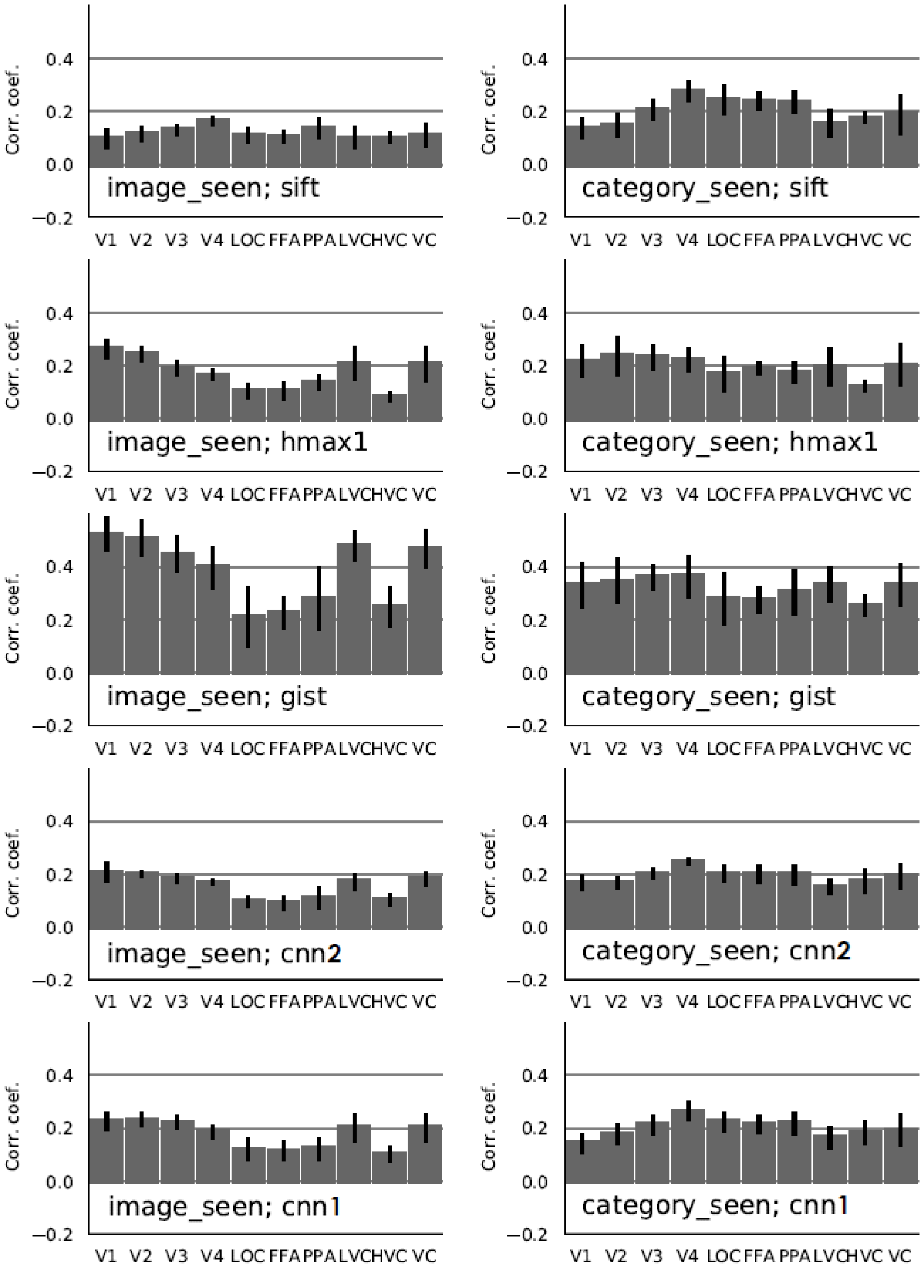

Figure 4.1: Correlation Coefficient

Chapter 5

Implementation & Result

5.1 Implementation

CNN, HMAX, GIST, SIFT+BoF; these 4 types of computational models were used to extract visual features from provided images. CNN and HMAX can mimic the hierarchy of a human's visual system. GIST was designed to recognize scenes. SIFT+BOF performed object recognition. These 4 models which were implemented in the data set are described below in detail.

5.1.1 CNN

CNN is a very effective model for image classification tasks. They are frequently used in analyzing visual imagery and also in working behind the scenes in image classification. In CNN, at first, the input image is sent through the convolution layer, then the max-pooling layer. These two stages perform the task of dimension reduction. These 2 stages can be applied multiple times until the image dimension becomes close to the filter dimension. MatConvNet implementation of the CNN model was applied. We used the images in ImageNet to train the CNN model in order to classify 1000 object categories. Due to computation difficulty, the number of layers was reduced to 2 from 8. 1000 units were randomly selected for these layers. Each image was represented by those unit's output vectors. The two layers were named CNN1 and CNN2.

5.1.2 HMAX

HMAX is a feedforward model that learns features hierarchically and performs recognition tasks. The combination of an image layer with six subsequent layers (S1, C1, S2, C2.S3, C3), creates these hierarchical layers and are built from previous layers by alternating template matching and max operations. We planned to use 3 layers of HMAX, but due to computational difficulties, we used 1 layer. Each image was represented by a vector of a single type of HMAX feature. We named it HMAX1.

5.1.3 GIST

GIST descriptors are a representation of a low-dimensional image that contains enough information to identify the scene in an image. In order to compute GIST, at first image conversion to greyscale was required. Then we made sure the max width did not exceed 256 pixels. Next, Gabor filters which has 16 orientations and 4 scales were used to filter the image. Then, the filtered image was segmented into a 4*4 grid. After that, filtered outputs in each block were averaged to extract 16 responses for each filter. Finally, we concentrated the responses received from multiple filters so that we could create feature vectors with 16 orientations, 4 scales and 16 block for each image.

5.1.4 SIFT with BoF

The scale-invariant feature transform (SIFT) is a feature detection algorithm in computer vision to detect and describe local features in images. SIFT descriptors were used to calculate the visual features using SIFT+BoF. SIFT descriptors were extracted from each image and were quantized into visual words using k-means clustering visual words. BoF histogram for each image was created from calculating the frequency of each visualwords. Finally, the histograms obtained through the above processing under- went L-1 normalization to become unit norm vectors.

5.2 Results

In this section the accuracy from the four models: CNN, HMAX, GIST and SIFT are compared.

Model Name	Maximum Accuracy(%)	Average Accuracy(%)
CNN1	85.54%(VC)	83.33%
CNN2	83.84%(VC)	80.81%
HMAX1	69.81%(V3)	66.77%
GIST	72.22%(VC)	71.69%
SIFT	80.07%(V4)	77.10%

Table 5.1: Results

5.3 Accuracy Comparison on Related Works

From Kamitani Lab, a paper published in nature, which achieved a stunning accuracy of 94% maximum. However, they used way more computational power to train with optimization in algorithm level. Also, they used multiple layers of different models to achieve the accuracy which we were unable to obtain.

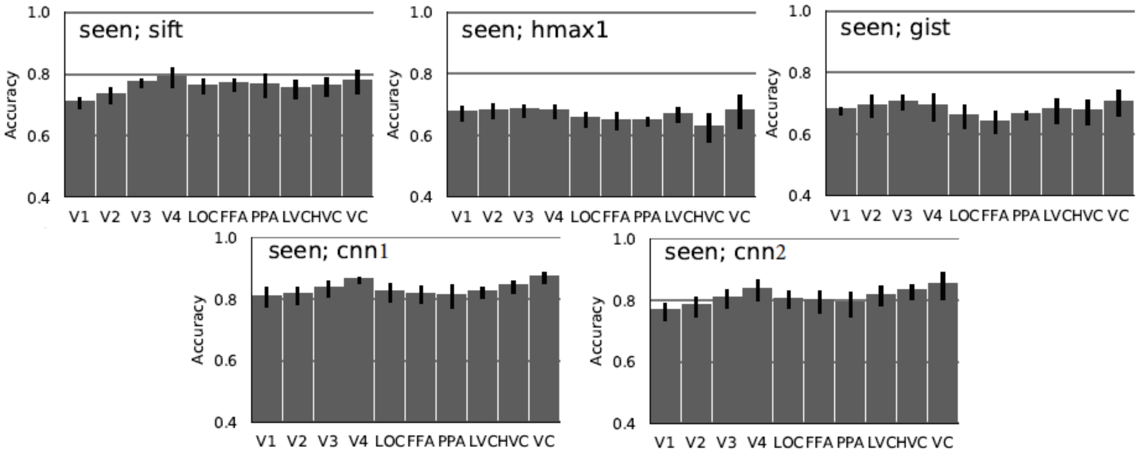

Figure 5.1: Accuracy of Our Model

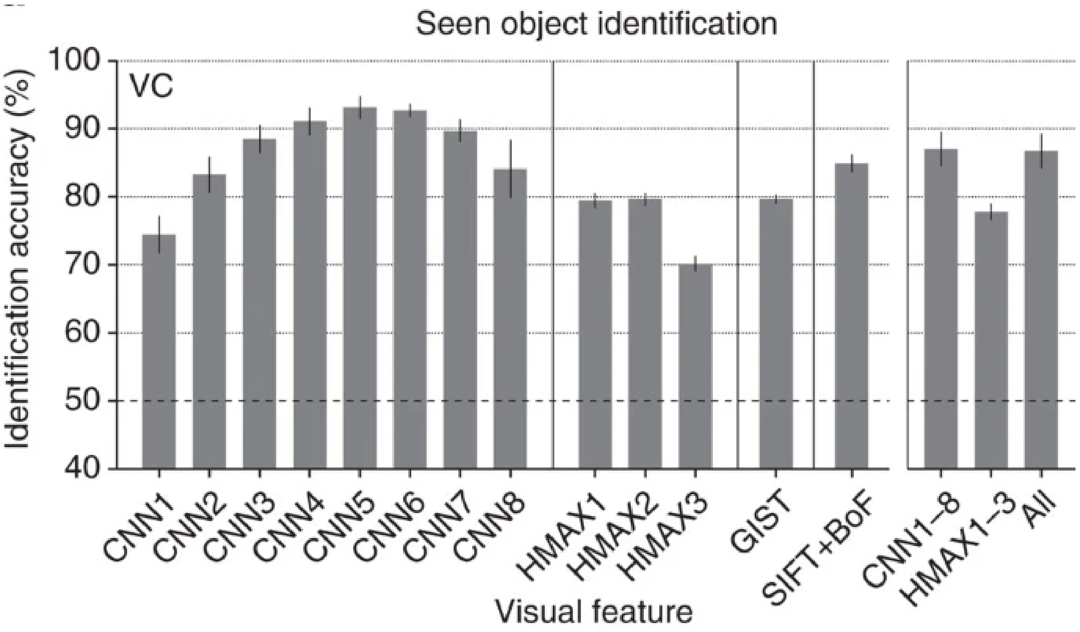

Figure 5.2: Accuracy of Previous Work

Chapter 6

Conclusion

6.1 Summary

In this modern era, humans and computers are co-operating in several fields. Since the last decade, we have successfully developed a large number of computing devices resulting in fewer risk and increased efficiency. Scientists and researchers throughout the globe have been spending their valuable time and energy to break this wall between a human and machine. However, the fundamental difference between a computer and a human still remains. A machine does not perceive objects and images the way human being does. For a machine, any real world object is just a bunch of signals, whereas for humans, it is way more complex. Humans do not have any frame rate, neither do they have time series data like computers have. On the other hand, computers do not possess intelligence to process and categorise real world object the way human being does. Due to that, we needed a generalized and a robust approach to tackle this issue. We have tried to provide such an approach in this paper. However, a lot more research need to be done in this field in order to bring it to the mainstream. It is hard to tell what technology can do until it comes on the hands of developers and researchers. As we were able to achieve some respectable results in the paper despite having some shortcomings, we are hopeful that it holds a lot of potentials to unlock.

6.2 Fields of Application

This research has a lot of potentials to be applied in the real world. Some of those are discussed below:

- **Facilitating disabled people:** Due to paralysis or brain disorder, a person may lose the ability to express the feelings to others. Sometimes, people get this disorder by born. Till now, the only practical way to communicate with them is to have an experienced person guess their feeling. However, this process is not fool proof and cumbersome. Our approach will open a door to tackle this issue by decoding their fMRI signals into human language.

- **Revisit human cognition:** Right now, human cognition is sort of a black box for us. It takes huge amount of efforts to just get started with human cognitive study. Also, a lot of human cognition is yet to be discovered and explored. Our solution will provide a way to revisit human cognition differently than before.

- **Crime Investigation:** Crime Investigation and Lie Detection are 2 crucial tasks for police. Most of these tasks are done through facial expressions, heart rate and blood pressure monitoring. However, people can learn to control and fake facial expressions and blood pressure to pass those tests. fMRI based solutions will be more robust than those approaches. Additionally, it can be used in conjunction with the previous approaches.

- **Providing Personalized Content:** In web, Content Personalization is a popular topic right now. From a video steaming to a e-commence website implements this feature in some shape or form. A lot of money and time get invested in implementing this feature. However, human behaviour is extremely complex. As a result, the potential contents for a customer cannot be predicted just by using their search or browsing history. Instead, their fMRI profile may give way more sophisticated insight than those previous approaches.

6.3 Future Improvements

- **Gain more domain knowledge** fMRI is an extremely sophisticated field of knowledge. Getting most out of fMRI data requires thorough domain knowledge in this field. As we did not have a very good grasp in this domain, some parts of the algorithm became extremely challenging to implement. Therefore, we need to gain more knowledge in order to improve the proposed model.

- **Train with more test subjects:** Our current model is trained with limited amount of test subjects. However, in order to generalize the model, we need to train against greater number of subjects. Our future goal would be to train with more test subjects.

- **Use distributed training with more powerful machines:** Training a model from fMRI data takes a lot of CPU power. Even with moderate CPU and 3 computing nodes, our training took more than 60 hours. It can be improved using more computing nodes. Modern distributed and parallel approach of training has improved, which can be used to train the model faster

- **Add imaginary object classifier:** We have to bear in mind that fMRI data does not only correspond to perceived objects through eyes, rather imaginary objects as well. Our current model lacks this feature. In future, this feature can be added in order to improve it.

- **Optimize the algorithm:** Right now, the algorithm lacks some of the performance tweaks for efficient model training. As a result, it takes a lot of time, which can be reduced by a significant amount of time by optimizing it.

Bibliography

[1] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. Tanabe, N. Sadato, and Y. Kamitani, "Visual image reconstruction from human brain activity using a combination of multiscale local image decoders," *Neuron*, vol. 60, pp. 915–929, Jan. 2009. DOI: 10.1016/j.neuron.2008.11.004. [Online]. Available: https://doi.org/10.1016/j.neuron.2008.11.004.

[2] T. Naselaris, R. Prenger, K. Kay, M. Oliver, and J. Gallant, "Bayesian reconstruction of natural images from human brain activity," *Neuron*, vol. 63, pp. 902–915, Sep. 2009. DOI: 10.1016/j.neuron.2009.09.006. [Online]. Available: https://doi.org/10.1016/j.neuron.2009.09.006.

[3] T. Naselaris, K. Kay, S. Nishimoto, and J. Gallant, "Encoding and decoding in fmri," *NeuroImage*, vol. 56, pp. 400–410, May 2011. DOI: 10.1016/j.neuroimage.2010.07.073. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2010.07.073.

[4] M. Johnson and M. Johnson, "Decoding individual natural scene representations during perception and imagery," *Frontiers in human neuroscience*, vol. 8, p. 59, Feb. 2014. DOI: 10.3389/fnhum.2014.00059. [Online]. Available: https://doi.org/10.3389/fnhum.2014.00059.

[5] T. Naselaris, C. Olman, D. Stansbury, K. Ugurbil, and J. Gallant, "A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes," *NeuroImage*, vol. 105, pp. 215–228, Oct. 2014, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2014.10.018. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2014.10.018.

[6] J. Han, X. Ji, X. Hu, L. Guo, and T. Liu, "Arousal recognition using audio-visual features and fmri-based brain response," *IEEE Transactions on Affective Computing*, vol. 6, no. 4, pp. 337–347, 2015. DOI: https://doi.org/10.1109/TAFFC.2015.2411280. [Online]. Available: https://ieeexplore.ieee.org/document/7056522.

[7] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature Communications*, vol. 8, no. 1, p. 15 037, May 2017, ISSN: 2041-1723. DOI: 10.1038/ncomms15037. [Online]. Available: https://doi.org/10.1038/ncomms15037.

[8] K. Han, H. Wen, J. Shi, K.-H. Lu, Y. Zhang, D. Fu, and Z. Liu, "Variational autoencoder: An unsupervised model for encoding and decoding fmri activity in visual cortex," *NeuroImage*, vol. 198, p. 136, 2019, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2019.05.039. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811919304318.

[9] M. N. I. Qureshi, J. Oh, and B. Lee, “3d-cnn based discrimination of schizophrenia using resting-state fmri,” *Artificial Intelligence in Medicine*, vol. 98, pp. 10–17, 2019, ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2019.06.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365719301393.

[10] M.-P. Hosseini, T. X. Tran, D. Pompili, K. Elisevich, and H. Soltanian-Zadeh, “Multimodal data analysis of epileptic eeg and rs-fmri via deep learning and edge computing,” *Artificial Intelligence in Medicine*, vol. 104, p. 101 813, 2020, ISSN: 0933-3657. DOI: https://doi.org/10.1016/j.artmed.2020.101813. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365718306882.