Gender Classification of Diffusion-Weighted Brain Images via Convolutional Neural Network

Spencer Diehl
bigdiehl@stanford.edu

Chandra Rajyam
rajyamch@stanford.edu

Martin Freeman
emmerich@stanford.edu

I. INTRODUCTION

The development of machine learning technologies is rapidly changing the methods by which 3D scans produced during magnetic resonance imaging (MRI) can be used to predict diagnoses. In this project we investigate the effectiveness of two machine learning approaches to classifying the diffusion tensor imaging (DTI) data of the Human Connectom Project (HCP) dataset by gender. Gender serves as a simple cornerstone to build and readily test various classification approaches on a relatively data-rich label. Given enough small developments in the application of learning to DTI data, it follows that other labels suggestive of a particular diagnosis may be successfully used in a classifier to earlier identify these diagnoses' in any subjects who receive a MRI. Thus, this project aims to evaluate the effectiveness of using DTI data to train a learning model to predict the gender of any given subject. Specifically, we will evaluate the effectiveness of both a classical learning approach and a deep learning approach utilizing a convolutional neural network.

The classical approach makes use of a common image processing algorithm, Scale Invariant Feature Transform (SIFT)[REF], to detect keypoints in the input images and resolve associated descriptors. A bag of visual words (BOVW) approach via k-means clustering is then used to cluster the feature space of the SIFT-processed training images into a desired number of $M = \text{(Number of Clusters)}$ representative features. The training data is then processed individually similar to the lumped training data to generate a set of M clustered features per example. A histogram containing the frequency of M overlapping features between the training example and the overall model is then input as a single datapoint into the training of a Support Vector Machine (SVM) that optimizes a decision boundary between the binary-labelled example.

The deep learning approach makes use of a neural network with convolutional layers to identify these feature descriptors instead, and is trained end-to-end on a single diffusion direction due to limitations in memory when handling data concatenation of different directions.

DTI is a special kind of diffusion weighted imaging (DWI) which uses specific MRI sequences to generate contrast in images by measuring the diffusion of water molecules. In DTI scans, the measurement of these diffusion rates have been used to map the white matter fibers that connect different parts of the brain. These images are essentially a 3-dimensional voxel array with elements corresponding to intensities of the diffusion gradient in a particular direction. Per subject image, we have $N = 30$ unique directions of this diffusion, and thus 30 input images associated with a particular label that can be consolidated as a single, concatenated example or taken as N individual examples.

In this project we aim to evaluate the effectiveness of using DTI data to predict the sex of test subjects within the dataset. It is hoped that this will provide insight into the best techniques that may be used to predict more advanced traits and actual diagnoses.

II. METHOD

A. Data Pre-Processing

The data from the HCP dataset is composed of DTI scans from 1200 subjects. The subjects were an approximately 50/50 mix of males and females who were 22-35 years of age. All subjects were deemed as healthy when participating in the data acquisition.

Each subject's DTI scan is composed of 288 diffusion weighted images, each image 145x174x145 in dimension. Each 3D image measures the water diffusivity at each voxel (i.e. 3D pixel) along a particular direction.

In order to facilitate digital transfer of the data, the DTI scans were compressed by a factor of 16 (500GB down to 30GB for the entire dataset). This was accomplished by down-sampling the scans to 30 images per subject, each image 64x64x64 in dimension. Visually, the images retain their general structure but fine details and features appear to be lost in the compression when compared to the raw images. Without the larger, raw dataset, we can't necessarily say whether or not the compression will have a significant effect on the performance, so the assumption is made that both the traditional approach and CNN approach will still be valid given the compression. Given this assumption, more weight is put toward gauging the relative performance between approaches rather than absolute performance.

B. Traditional Approach

1) Feature detection: As our data consists of 3D diffusion-weighted images, manual selection of label-indicative keypoints for each image was not found to be an efficient approach. Automated feature detection in our dataset was

desired, and based on our criteria of capturing robust and adequately complex features, the Scale-Invariant Feature Transform (SIFT) algorithm was chosen. In addition to scale invariance, which proves useful after observing that our images are not necessarily normalized for scale, SIFT ensures features are rotation and viewpoint invariant as well. Any amount of rotation and viewpoint invariance enhances the quality of the detected features over the entire dataset as we have multiple principal directions for a given subject, allowing conflation of homologous features across these principal directions that would have otherwise been dissociated from one another.

One obvious hindrance in applying the SIFT algorithm is the traditional implementation for SIFT uses 2D images. SIFT3D, an analogue of the SIFT algorithm accessible via a cross-platform C library, was leveraged to extend SIFT into a 3D imagespace and automatically generate keypoints and their corresponding descriptors for a given image [1].

2) K-means clustering, SVM and dimensionality reduction: A bag of visual words (BOVW) approach is used to compartmentalize the feature-space of the dataset, making use of k-means clustering to consolidate similar descriptors. While this does not help us reduce dimensionality of the features, it does help us by developing a smaller set of feature descriptors that adequately represent the structure of the feature-space. By performing this clustering on our training set, we develop a model for which we can fit a new image to by observing which cluster "bins" their descriptors best fall into.

Given our model (a set of clustered features over the entire training set), we take each individual training image and process it through individual k-means clustering to generate an identical number of clustered features to our model. From this, a given training image's distribution of clustered features is fit to the model (the overall training set's distribution of clustered features) by generating a histogram of overlapping features and their frequency. The histogram for the image, as well as the associated label (male or female), are fed as a singular datapoint to a SVM, where after all the training histograms are generated, the decision boundary between distributions typical to each label is optimized.

As the dimensionality of the histogram is equal to the number of clusters we select in the lumped k-means clustering step, using larger cluster sizes leads to longer fitting time for the SVM. Opting to reduce this fitting time, we incorporated the use of principal component analysis (PCA) on the set of training histograms to reduce dimensionality for SVM processing (400 clusters down to 15) and generate a set of representative linear, uncorrelated variables that compactly describe each histogram's distribution

C. Convolutional Neural Network

As a baseline for using a convolutional neural network (CNN), only a single diffusion direction was used to train a CNN designed to process 3D images. The architecture for the CNN was developed using Keras and the Tensorflow backend, The encoding layer was composed of four convolution layers,

each layer performing a 3D convolution, a batch normalization, and a 3D pooling procedure. The prediction layer was composed of three fully connected neural network layers. The final layer consisted of a single node that used a sigmoid activation function to output a probability value.

In each epoch of training, a random batch of images (about half the dataset) was used to train the CNN. The batch divided into sets containing the left and right brain hemispheres. The CNN was then sequentially trained on these two sets. To monitor the progress, every 10 epochs the trained CNN was evaluated on a testing set that was also divided into two sets containing the left and right hemispheres.

Due to the relatively small subject sample size, the training dataset was artificially augmented by including randomly rotated and translated versions of images randomly chosen from the original input dataset. This was done to not only provide more training examples, but also to help the model more robustly generalize to new images.

In order to maximize the use of the limited data, k-fold cross validation was implemented to evaluate the effectiveness of the trained model.

III. PRELIMINARY EXPERIMENTS

A. Traditional Approach

Testing the classical pipeline on a single diffusion direction produced a poor average prediction accuracy. The trained model achieved an average accuracy of 60% during a 5-fold cross validation procedure across a multitude of different parameters (cluster sizes, PCA reductions, SIFT parameters). While we expected to achieve greater performance with this traditional pipeline, a multitude of justifications arise in retrospect that grant us some insight into this poor accuracy. Namely, the anatomical differences between male and female brains lay primarily in certain sections of the brain having slight volume differences. Differences in volume, or scale, as opposed to structure and intensity, aren't easily detectable features by the selected SIFT algorithm. In fact, the scale invariance of SIFT likely reduced the quality of detected features.

Another consideration is the effect of major compression of the images in the dataset. Undoubtedly, if there are fine features indicative of a particular gender or age label, their quality is washed out with compression.

Lastly, this traditional approach was trained over only one direction on a 200-subject subset of the overall data due to the large computation time. With consistent 60% performance over just this subset of the data, we could see a moderate improvement in prediction accuracy given training on the full data set.

B. Convolutional Neural Network

Testing the CNN on a single diffusion direction produced promising initial results. The trained classifier was able to achieve an average of 90% prediction success rate during a 5-fold cross validation procedure.

IV. NEXT STEPS

Ultimately, our project objective is to develop an accurate, robust CNN for DTI classification. We explored a traditional pipeline as a means to develop a baseline performance we can measure the CNN against, and while there may be other pipelines that employ traditional machine learning approaches that may perform better, we intend to take what we've learned from the approach we used, particularly, lessons pertaining to the feature representation of anatomical differences between brains in different age and gender categories outlined in the literature[2,3], as we devote the rest of the project to developing and testing the CNN.

For the CNN, the next step would to be train the baseline CNN across each of the 30 directions. From these CNN models, intermediate features across different directions can be concatenated and an overall predictor can be built on top of these features.

In addition, it may be possible to determine which areas of the brain are most useful in predicting the sex of the subjects. A next step would then be to identify these regions and focus the training of the CNN's on these "patches" in the images. Focusing on subsets of the images would also make it more feasible to use higher resolution data when training and testing.

As the models are developed, training to predict for other traits (such as age) may be explored, which will be useful to determine how these learning methods may be extended beyond sex prediction.

CONTRIBUTIONS

While team members generally worked across many aspects of the project, specific contributions are as follows:

Spencer Diehl: Contributed to CNN design and helped determine computing requirements.

Chandra Rajyam: Contributed to traditional pipeline approach.

Martin Freeman: Contributed to traditional pipeline approach.

CODE

The Github repo containing the code used in the project can be found at:

https://github.com/chandra-rajyam/cs229project

REFERENCES

[1] [1] B. Rister, M. A. Horowitz and D. L. Rubin, "Volumetric Image Registration From Invariant Keypoints," in IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4900-4910, Oct. 2017. doi: 10.1109/TIP.2017.2722689

[2] [2] D. D. Franks, Sex Differences in the Human Brain, Oxford Handbooks Online, Jul. 2018.

[3] [3] Ruigrok AN, Salimi-Khorshidi G, Lai M.C, Baron-Cohen S, Lombardo M.V, Tait R.J, Suckling J. 2014. A meta-analysis of sex differences in human brain structure. Neurosci Biobehav Rev. 39:3450.