EVALUATING DIMENSIONALITY REDUCTION TECHNIQUES FOR FEATURE GENERATION FROM PARKISONS' BRAIN IMAGING

by

RAJESWARI SIVAKUMAR

(under Direction of Shannon Quinn)

ABSTRACT

Parkinson's disease is one of the most prevalent neurodegenerative diseases, second only to Alzheimer's. There is an urgent need to improve detection through means outside observation of motor symptoms. The present research is to examine diffusion tensor images (DTI) from Parkinson's and control patients through linear dynamical systems and tensor decomposition methods to generate features for training classification models. We will reduce the dimensionality of these images to allow us to focus on the key features that differentiate PD and control patients. We show through our experiments that these approaches can result in good classification accuracy (90%) and indicate this avenue of research has a promising future. In our second experiment we further examine the effect of orientation on features. We find that we are able to produce robust features regardless of orientation. We are able to train a classifier with these features to get up to 94% accuracy.

INDEX WORDS: Parkinson's Disease, Biomedical Data Analysis, Machine Learning, Diffusion Tensor Imaging, Image Analysis, Feature Extraction, Singular Value Decomposition, Tensor Decomposition, Dimensionality Reduction

EVALUATING DIMENSIONALITY REDUCTION TECHNIQUES FOR FEATURE GENERATION FROM PARKISONS' BRAIN IMAGING

by

RAJESWARI SIVAKUMAR

B. A. University of California, Irvine, 2014

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2019

© 2019

Rajeswari Sivakumar

All Rights Reserved

EVALUATING DIMENSIONALITY REDUCTION TECHNIQUES FOR FEATURE GENERATION FROM PARKISONS' BRAIN IMAGING

by

RAJESWARI SIVAKUMAR

Major Professor: Shannon Quinn

Committee: Khaled Rasheed

Frederick Maier

Electronic Version Approved

Ron Walcott

Interim Dean of Graduate School

December 2019

TABLE OF CONTENTS

LIST OF TABLES ..... *v*
LIST OF FIGURES ..... *vi*
CHAPTER 1 ..... *1*
INTRODUCTION AND LITERATURE REVIEW 1
Understanding the progression of Parkinson's Disease 1
Brain Imaging Techniques 2
Previous Work 4
Tensor and Matrix Decomposition 5
Autoregressive Techniques 7
CHAPTER 2 ..... *9*
PARKINSON'S CLASSIFICATION & FEATURE EXTRACTION FROM DIFFUSION TENSOR IMAGES 9
Introduction 10
Parkinson's Disease 10
Parkinson's Progression Markers Initiative Datasets 11
Related Work 11
Tensor and Matrix Decomposition 13
Methods 15
Algorithm Selection 16
Experiment 1 16
Experiment 2 17
Results 17
Experiment 1 17
Experiment 2 18
Discussion 20
References 20
CHAPTER 3 ..... *24*
ANALYZING LDS FEATURES AS A FUNCTION OF ORIENTATION 24
Introduction 25
Methods 26
Results 28
Discussion 31
References 32
CHAPTER 4 ..... *33*
CONCLUSION 33
REFERENCES *34*

LIST OF TABLES

Table 1: Classification accuracy of features generated from Tucker decomposition after various additional dimensionality reduction techniques are applied 18
Table 2: Classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied..... 19
Table 3: Cross-validated(n=5) classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied. (Orientation 1, coronal slices) 29
Table 4: Cross-validated(n=5) classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied. (Orientation 2, sagittal slices)..... 29
Table 5: Cross-validated(n=5) classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied. (Orientation 3, axial slices) 30
Table 6: Mean Welch's T-statistics for Parkinson's and control groups when comparing features from different slice orientations. 30
Table 7: Resulting p-value corresponding to mean t-statistic comparing orientation as function of patient condition 31

LIST OF FIGURES

Figure 1: Tucker decomposition, visualized**.** Similar to SVD, it is used to compress tensors. We are thus able to use it as means to describe brain images without breaking down specific regions of interest or or focusing on specific brain images. 15
Figure 2: a) (left): Slice from original brain image at a specific time point; b) (right): Corresponding slice from tensor decomposition output..... 19
Figure 3: Illustration to show how slices are taken from the brain image for the different orientation conditions. Axial refers to vertical slices taken front to back. Coronal refers to horizontal slices. Saggital slices are vertically taken left to right. 27

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Understanding the progression of Parkinson's Disease

Parkinson's disease (PD) is complex and affects multiple parts of the human nervous system, but the various stages of it's progression up the central nervous system (CNS) can be related back to one key factor, the α-synuclein protein. This protein is naturally occurring in many different types of neural cells, and there must be a large amount of this protein present for Parkinson's to manifest. In normal healthy cells, α-synuclein floats freely in cytosol concentrated mostly in the neuronal somata and terminal buotons. It is typically hydrophilic and has a strong affinity for membranes of cell vessicles. In Parkinson's patients this protein can become denatured and lose hydrophility. In this event, it tends to form β-sheets that tend to clump together. Only in Parkinson's does this occur. This denaturation of α-synuclein is not a part of the normal aging process nor does it occur in other neurodegenerative diseases. (Braak et al., 2004) Once aggregates of the α-synuclein protein begin to form, they will eventually develop into pale, branching inclusion neurites or Lewy neurites (LN) . This pathology will persist eventually forming, round pale granular aggregations known as Lewy bodies or inclusion bodies. (Del Tredici and Braak, 2016)

The α-synuclein protein is also interesting in that only specific nerve cells are susceptible to the Lewy pathology (LP) that it causes, regardless of location in the brain. Projection cells with long, axons that are not well myelinated are the most vulnerable to α-synuclein aggregates. Conversely neurons with short or strongly myelinated axons are largely unaffected. As a consequence,

Parkinson's develops and spreads in the CNS in consistent, well documented patterns. (Del Tredici and Braak, 2016)

This spread and development of PD can be divided into 6 stages of increasing severity, with stages 1-3 corresponding to pre-symptomatic pathology and stages 4-6 corresponding to symptomatic pathology. It typically begins in the olfactory regions of the central nervous system and moves upward. By stage 4, LBs can be seen in leading from the dorsal motor nucleus to the substantia nigra. Around this stage of development in the pathology, onset of motor symptoms can start to be noticed and lesions in the anteromedial temporal cortex. Decline in this region marks the beginning of cognitive decline and memory issues as it is an important pathway of communication between the prefrontal cortex and limbic system. As the disease progress, it eventually climbs up the central nervous system, taking over the neocortex sometimes reaching even the premotor and sometimes primary systems. (Braak et al., 2004)

Brain Imaging Techniques

There are a multitude of brain imaging techniques to measure structure and function of the brain. The earliest of these is electroencephalography or EEG. It involves placing several electrodes on the scalp and measuring post-synaptic electrical activity from neurons. The key type of EEG studies are those involving the examination of event related potentials or ERPs. These measure changes in electrical activity following the presentation of specific stimuli. The premise of these studies is to compare differences in the magnitude of ERPs across different subject groups. It is excellent for this type of task as EEG has excellent temporal resolution, however it is less suited to localizing response as it significantly lacks in spatial resolutions (Dale and Halgren, 2001).

Furthermore, as functional changes (in behavior) don't tend to occur until much later in the disease timeline, it is less useful to focus on neurological responses.

To remedy this, various forms of brain imaging that focus on studying structure, have been developed over the years. Key among these is magnetic resonance imaging or MRI. This relies on using a magnetic field and changing the spin of hydrogen molecules in tissue. Once the field is deactivated the molecules return to their resting state at varying rates depending on the type of tissue they are present in. In this manner by tracking this change we are able to construct three-dimensional (3D) images of the human brain and body.

Other brain imaging modalities that focus on the development of features include single photon emission computed tomography (SPECT) and positron emission tomography (PET). Both of these use similar techniques to image tissue as a radioactive tracer emits gamma radiation or positrons respectively (Lewine, et al., 2007). A subset of SPECT, known as DatScan, uses the tracer Ioflupane/FIP-CIT-I-123 to evaluate uptake of dopamine, which is known to be reduced in Parkinson's patients (Martinez-Murcia et al., 2013). While SPECT and PET both provide good spatial resolution, they are extremely invasive due to the use of radioactive tracers. For this same reason, they may rule out certain individuals prone to adverse reactions from the radioactive tracers. MRI for this reason becomes the favored choice, with the ability to provide good contrast images without these drawbacks.

A specific subset of MRI, diffusion tensor imaging (DTI), tracks diffusion of water molecules to capture microstructural changes (Soares et al., 2013). In general the diffusion of water in biological materials is restricted or hindered by cell membranes. With necrosis of tissue, diffusivity in and out of the cell increases. For this reason, DTI is particularly good for identifying

minute changes in the central nervous system. DTI produces a diffusion matrix at each voxel to indicate magnitude and direction of diffusion. When visualizing, these are mapped to scalar values that can be used to generate a pixel array. We aim to examine these scalar maps for feature extraction.

Previous Work

A variety of tools currently exist for diagnosis of Parkinson's through pre-motor symptoms. For example Parkinson's seems to measurably affect olfactory sensitivity prior to presenting motor symptoms more than other motor neuron diseases, as illustrated by the University of Pennsylvania Smell Identification Test (UPSIT). While there is still more work needed to refine tests like these, it is one example that proves the feasibility of earlier diagnosis of Parkinson's disease.

The PPMI was established under the premise that there are key biomarkers that can be used to better observe and diagnose Parkinsons. Chahine & Stern, (2016) through evaluation of previous research determined that several biomarkers can be studied to better understand Parkinson's. Among these are behavioral observations of motor function, accelerometer-based observation of motor skill, biofluids, peripheral tissue, imaging, genetics. Many researchers have sought to quantitatively evaluate these biomarkers to describe and classify Parkinson's from biomarkers alone.

On group of researchers, (Adeli et al., 2017) achieved notable results using kernel-based features to train a classification framework. The kernel-based selection approach allowed the researchers to select features best suited for different classification algorithms. Accuracy of 70.5% and 95.6% were obtained for features found from MRI and SPECT data respectively.

Other researchers, Banerjee, et al. (2016) were able to achieve 98.53% using ensemble learning methods trained on T1 weighted MRI data. However, Banerjee, et al. used several domain knowledge based feature extraction methods to preprocess their data including image registration, segmentation, and volumetric analysis.

Tensor and Matrix Decomposition

Matrix decomposition has been used in a variety of computer vision applications in recent years including analysis of facial features. It offers a another means of quantifying the features that describe the relationships between values in a 2D space and can be generalized to a variety of applications. The key being that decomposition offers a powerful means of simultaneously evaluating the relationships of values in a 2 or higher dimensional space. In higher dimensional spaces, tensor decomposition is used, where tensors are a generalization of matrices. (Rabanser, et al. 2017)

Matrix decomposition can be described as a means of separating a matrix into several component matrices whose product would result in the original matrix. For example when solving a system of equations you might approach formulate the problem as:

$$Ax = b,$$

where $\boldsymbol{A}$ is a matrix and $\boldsymbol{x}$ and $\boldsymbol{b}$ are vectors. When trying to solve this equation, we could apply a matrix decompositions operations to the matrix $\boldsymbol{A}$, to more efficiently solve the system. By finding the products of the of $\boldsymbol{x}$ and $\boldsymbol{b}$ with the the one matrix resulting from the decomposition and the inverse of the other, we can solve the system of equations with significantly fewer operations.

We can generalize this premise to machine learning, when model complexity of models, often result in exponential increases in number of computations. This also affects the applications of new algorithms and pipelines can be used in because of their complexity.

We can choose specific types of decompositions that also allow us to preserve unique information about original matrix while also reducing the the size of the matrix. For example, in the case of singular value decomposition we are trying to solve:

$$A = USV^T,$$

Where A is the original matrix, of size $m \times n$, U is an orthogonal matrix of size $m \times n$, S is a diagonal matrix of size $n \times n$, and V^T is an orthogonal matrix of size $n \times n$. This generalization of the eigendecomposition is useful in compressing matrices without losing information. It will come into play with our final experiment using linear dynamical systems to extract features from the DTIs.

Extending the premise of singular value decomposition (SVD) to higher order matrices, or tensors, we come to Tucker decomposition. Tucker decomposition can be described by the equation, for an example of a third of:

$$T_1(i,j,k) = \sum_{r_1, r_2, =1}^{R_1,R_2,R_3} T_2\,(r_1,\ r_2, r_3) \cdot A(i,r_1) \cdot B(i,r_2) \cdot C(j,r_3)\ ,$$

Where T_1 is the original tensor of size $i \times j \times k$, and T_2 is the resulting, third order core tensor of size $R_1 \times R_2 \times R_3$. The tensors (which in this case are matrices) A, B, and C of size $i \times R_1$, $j \times R_2$, and $k \times R_3$ respectively (Rabanser, 2017).

Tensor decomposition has already been shown to be useful for feature extraction, as Phan and Chichocki (2010) illustrate. Using several processes similar to the one described above, they are able to successfully train a support vector machine (SVM) to classify features extracted from the MNIST handwritten digits dataset with accuracy of over 99 %.

Additionally this process is of interest to us because of its success in compression algorithms. For example Ruiters and Klein (2009) compare sparse tensor decomposition to other popular data compression algorithms. They are able to achieve compression ratios 3 to 4 times higher than state of the art methods, and preserve many high frequency textural details of the images using tensor decomposition. Because of this, we believe that we can productively train features obtained from using tensor decomposition and correlate our results to the key features in the original brain images.

Autoregressive Techniques

Autoregressive techniques have been previously used in dynamic texture analysis of videos to model. Specifically we focus on state-space models also known as linear dynamical systems (LDS). In this technique we are trying to find the appearance model

$$y_t = Cx_t + u_t$$

and the state model

$$x_t = Ax_{t-1} + Wv_t \ .$$

In the state model we are effectively trying to predict the next image in a sequence based on the previous, or a version of a moving average. This is however rather computationally inefficient do

do on whole images, thus our appearance model comes in, to make the problem smaller. You can think of y_t as an image flattened into a vector and C is an output matrix corresponding the low dimensional state space vector, x_t. We account for noise with u_t. The state space vector, x_t is found at each time point using singular value decomposition (SVD) to find the dot product on the top q singular values and columns of $\hat{V}^T$. We can then calculate the state-transition matrix A to which can then be applied to predict the next state in the model. (Hyndman, et al., 2007)

This method is selected because it will translate well to modeling the average change over the course of the timesteps of a diffusion tensor image.

CHAPTER 2

PARKINSON'S CLASSIFICATION & FEATURE EXTRACTION FROM DIFFUSION TENSOR IMAGES[1]

[1] Sivakumar, R and Quinn, S. Accepted by Scipy Conference 2019, Reprinted here with permission of the publisher.

Introduction

Parkinson's disease (PD) affects over 6.2 million people around the world. Despite its prevalence, there is still no cure, and diagnostic methods are extremely subjective, relying on observation of physical motor symptoms and response to treatment protocols. Other neurodegenerative diseases can manifest similar motor symptoms and often too much neuronal damage has occurred before motor symptoms can be observed. The goal of our study is to examine diffusion tensor images from Parkinson's and control patients through linear dynamical systems and tensor decomposition methods to generate features for training classification models. Diffusion tensor imaging emphasizes the spread and density of white matter in the brain. We will reduce the dimensionality of these images to allow us to focus on the key features that differentiate PD and control patients. We show through our experiments that these approaches can result in good classification accuracy (90%) and indicate this avenue of research has a promising future.

Parkinson's Disease

Parkinson's disease (PD) is one of the most common neurodegenerative disorders (Chaudhuri, et al., 2016). The disease mainly affects the motor systems and its symptoms can include shaking, slowness of movement, and reduced fine motor skills (Sveinbjornsdottir, 2016). As of 2015 an estimated 6.2 million globally were afflicted with the disease (Rabansar, et al., 2017). Its cause is largely unknown and there are some treatments available, but no cure has yet been found. Early diagnosis of PD is a topic of keen interest to diagnosticians and researchers alike. Currently Parkinson's is diagnosed based on the presence of observable motor symptoms and change in symptoms in response to medications that target dopaminergic receptors such as Levdopa. The problem with this approach is that it relies on treating symptoms instead of preventing them. Once motor symptoms present, at least 60% of neurons have been affected and there is little likelihood

of healing them fully. Additionally early diagnosis will help reduce likelihood of misdiagnosis with other motor neuron diseases.

Parkinson's Progression Markers Initiative Datasets

The Parkinson's Progression Markers Initiative (PPMI) is an observational clinical study designed to identify PD biomarkers [4] and contribute towards new and better treatments for the disease. The cohort consists of approximately 400 de novo, untreated PD subjects and 200 healthy subjects followed longitudinally for clinical, imaging and biospecimen biomarker assessment. The PPMI data set is a collection of biomarker data collected from a longitudinal study of Parkinson's and control subjects. They have thus far collected DaT scan, MRI, fMRI, and CT scan data from several hundred subjects in 6 month intervals. The first began collecting data in 2010, funded by the Michael J.Fox Foundation.

The dataset chosen for this paper was PPMI's Diffusion Tensor Imaging (DTI) records. DTI has been shown to be a promising biomarker in Parkinsonian symptoms [5] and can provide unique insights into brain network connectivity. Moreover, the DTI data was one of PPMI's cleanest and largest datasets and thus expected to be one of the most useful for further analysis. A DTI record is a four-dimensional dataset comprised of a time-series of a three-dimensional imaging sequence of the brain. PPMI's DTIs generally consisted of 65 time slices, each taken approximately five seconds apart.

Related Work

A variety of tools currently exist for diagnosis of Parkinson's through pre-motor symptoms. For example Parkinson's seems to measurably affect olfactory sensitivity prior to presenting motor symptoms more than other motor neuron diseases, as illustrated by the University of Pennsylvania

Smell Identification Test (UPSIT). While there is still more work needed to refine tests like these, it is one example that proves the feasibility of earlier diagnosis of Parkinson's disease.

The PPMI was established under the premise that there are key biomarkers that can be used to better observe and diagnose Parkinsons. Chahine & Stern, (2016) through evaluation of previous research determined that several biomarkers can be studied to better understand Parkinson's. Among these are behavioral observations of motor function, accelerometer based observation of motor skill, biofluids, peripheral tissue, imaging, genetics. Many researchers have sought to quantitatively evaluate these biomarkers to describe and classify Parkinson's from biomarkers alone.

On group of researchers, (Adeli et al., 2017) achieved notable results using kernel-based features to train a classification framework. The kernel-based selection approach allowed the researchers to select features best suited for different classification algorithms. Accuracy of 70.5% and 95.6% were obtained for features found from MRI and SPECT data respectively.

Other researchers, Banerjee, et al.(2016) were able to achieve 98.53% using ensemble learning methods trained on T1 weighted MRI data. However Banerjee used several domain knowledge based feature extraction methods to preprocess their data including image registration, segmentation, and volumetric analysis.

Our present research strikes a balance between the two. While our autoregressive model does utilize a basic understanding of relevance of time in diffusion tensor imaging, we do not utilize any other domain specific knowledge to inform our feature extraction. Our hope is to build a generalizable approach that can be applied to other data structured similarly both within and outside the domain of biomedical image analysis. Additionally we want to improve the models

being trained without domain specific knowledge on MRI data. This is because MRI is a far less invasive brain imaging method than SPECT imaging which is an X-ray based technique and must be used at a limited frequency. Additionally the multiple MRI modalities offer versatility in examining biological structures.

Tensor and Matrix Decomposition

Matrix decomposition has been used in a variety of computer vision applications in recent years including analysis of facial features. It offers a another means of quantifying the features that describe the relationships between values in a 2D space and can be generalized to a variety of applications. The key being that decomposition offers a powerful means of simultaneously evaluating the relationships of values in a 2 or higher dimensional space. In higher dimensional spaces, tensor decomposition is used, where tensors are a generalization of matrices.

Matrix decomposition can be described as a means of separating a matrix into several component matrices whose product would result in the original matrix. For example when solving a system of equations you might approach formulate the problem as:

$$Ax = b,$$

where $\boldsymbol{A}$ is a matrix and $\boldsymbol{x}$ and $\boldsymbol{b}$ are vectors. When trying to solve this equation, we could apply a matrix decompositions operations to the matrix $\boldsymbol{A}$, to more efficiently solve the system. By finding the products of the of $\boldsymbol{x}$ and $\boldsymbol{b}$ with the the one matrix resulting from the decomposition and the inverse of the other, we can solve the system of equations with significantly fewer operations.(Rabanser, et al. 2017)

We can generalize this premise to machine learning, when model complexity of models, often result in exponential increases in number of computations. This also affects the applications of new algorithms and pipelines can be used in because of their complexity.

We can choose specific types of decompositions that also allow us to preserve unique information about original matrix while also reducing the size of the matrix (Phan and Chichocki, 2010). For example, in the case of singular value decomposition we are trying to solve:

$$A = USV^T,$$

Where $\boldsymbol{A}$ is the original matrix, of size $\boldsymbol{m \times n}$, $\boldsymbol{U}$ is an orthogonal matrix of size $\boldsymbol{m \times n}$, $\boldsymbol{S}$ is a diagonal matrix of size $\boldsymbol{n \times n}$, and $\boldsymbol{V^T}$ is an orthogonal matrix of size $\boldsymbol{n \times n}$. This generalization of the eigendecomposition is useful in compressing matrices without losing information. It will come into play with our final experiment using linear dynamical systems to extract features from the DTIs.

Extending the premise of singular value decomposition (SVD) to higher order matrices, or tensors, we come to Tucker decomposition. Tucker decomposition can be described by the equation, for an example of a third of:

$$T_1(i,j,k) = \sum_{r_1, r_2, =1}^{R_1,R_2,R_3} T_2\,(r_1,\ r_2, r_3) \cdot A(i, r_1) \cdot B(i, r_2) \cdot C(j, r_3)\ ,$$

Where $\boldsymbol{T_1}$ is the original tensor of size $\boldsymbol{i \times j \times k}$, and $\boldsymbol{T_2}$ is the resulting, third order core tensor of size $\boldsymbol{R_1 \times R_2 \times R_3}$. The tensors (which in this case are matrices) $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ of size $\boldsymbol{i \times R_1}$, $\boldsymbol{j \times R_2}$, and $\boldsymbol{k \times R_3}$ respectively. (Rabanser, et al., 2017)

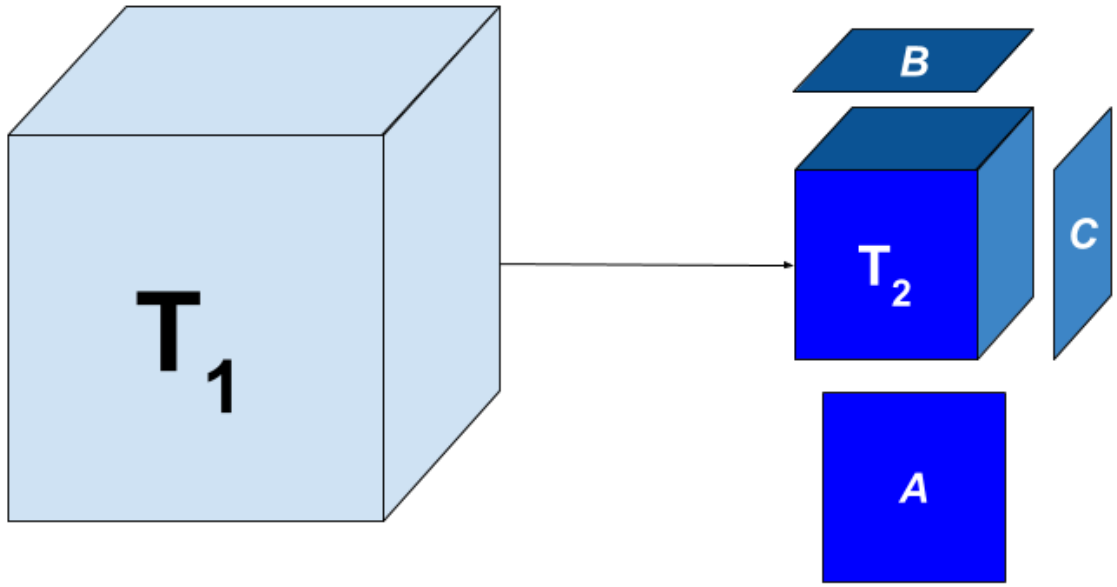

Figure 1: Tucker decomposition, visualized**.** Similar to SVD, it is used to compress tensors. We are thus able to use it as means to describe brain images without breaking down specific regions of interest or or focusing on specific brain images.

Methods

There are two main experiments conducted. We examine both Tucker tensor decomposition and a linear dynamical systems approach to reduce number of dimensions and scale down diffusion tensor images. The goal is to evaluate the two approaches for the quality of features extracted. To this end, the final feature vectors produced by each method is then passed on to a random forest classifier, where the accuracy of the final trained model is measured on a classification task to predict control or Parkinson's (PD) group.

Algorithm Selection

To guide our selection of a classifier, we used the python package TPOT (Olson, et al., 2016). TPOT uses genetic algorithms to iteratively generate, select and evaluate classification pipelines. We evaluated 10 generations of pipelines with population size 100 in each and found that Random Forest classification was most successful as predicting Parkinson's from the generated features.

Given the success of random forest classifier, we considered that we might further improve our accuracy by reducing the number of features we used from the generated set. We considered that because we are focused on the differences in a relatively small specific brain regions, only a small number of features would be relevant. To test this theory, we used three different methods to reduce the dimensionality of our feature set to 20 components: linear principle component analysis (PCA), linear discriminant analysis (LDA) and kernel PCA using a radial basis function (RBF).

Experiment 1

Using the tensorly package (Kossaifi, et al., 2016), a Tucker decomposition was applied to each brain image. This approach to tensor decomposition was selected because it produces one core tensor that is representative but scaled down from the original diffusion tensor image. Additionally Tucker decomposition, unlike other forms of tensor decomposition is significantly better at preserving features specific to the tensor being decomposed. Because of this it has applications in compression algorithms.

The Tucker decomposition method was chosen in the over other tensor decomposition methods to preserve features unique to each brain image it is applied to. This allowed us to scale down each image and focus features and regions of interest in each that are specific to that image.

In this experiment we decompose each brain image from a dimension of (65,100,116,116) to (10,10,10,10) to have a continuity in number of features produced.

Experiment 2

This experiment focused on breaking down the feature extraction further and evaluate another approach: linear dynamical systems. We scale down each coronal slice in the images and then evaluate the change over time. The reason for scaling down the coronal slices is to allow us to more efficiently build a transition model to represent the flow of water over the time steps of the image. This will allow us to build a three-dimensional representation of the brain from the images that will show the flow of water and the distribution of white matter in the brain. We evaluate the produced transition matrix as features to be applied to the classification pipeline.

Results

Experiment 1

While we were able to successfully classify images as belonging to PD or control subjects with an accuracy of 94% immediately (Table: 1), we were not able to improve on this by further reducing the produced features with various dimensionality reduction methods. In fact it appears that in some cases, such as linear discriminant analysis (LDA), additional dimensionality reduction adversely affects classifier performance. In exploring a slice of the output core tensor at one ‘time’ point, what we see suggests that the output of the tensor decomposition might be likened to a stack of sliced that focus on the regions of interest in the original image. This is validated by examining several corresponding decomposed core and original slices.

Table 1: Classification accuracy of features generated from Tucker decomposition after various additional dimensionality reduction techniques are applied

Dimensionality Reduction Method	F-measure	Accuracy
-	0.94	0.94
PCA	0.94	0.94
LDA	0.82	0.81
Kernel PCA	0.94	0.94

Experiment 2

We were able to achieve accuracy of 82% with random forest classifier alone. This outperforms previous benchmarks in training classifiers on synthetic features derived from MR images (Cole, et al., 2016). Compared to present results Cole et al. (2016) achieved only 70% accuracy at best on synthetic features generated from T1 weighted MRI scans. Furthermore, based on the F-measure scores across the experiment conditions, we can reasonably say that our model is not skewed as a consequence of the uneven distribution of the data.(Table: 2) The PPMI data is heavily skewed toward Parkinson's individuals, with a majority of our data set coming from Parkinson's patients (421 subjects) versus controls (213 subjects), which was also addressed by rebalancing the classes by oversampling the control.

We intuited that we could speed up model training and improve accuracy by reducing the number of synthetic features we retained. We initially tried linear PCA and LDA to perform the dimensionality reduction. However, these actually hurt performance, resulting in test accuracy of 81% and 74% respectively. Based on this, we considered non-linear dimensionality reduction

would be more effective. To this end we used Kernel PCA with RBF kernel, which effectively improved accuracy to 89%.

Table 2: Classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied.

Dimensionality Reduction Method	**F-measure**	**Accuracy**
-	0.90	0.82
PCA	0.89	0.81
LDA	0.84	0.74
Kernel PCA	0.93	0.89

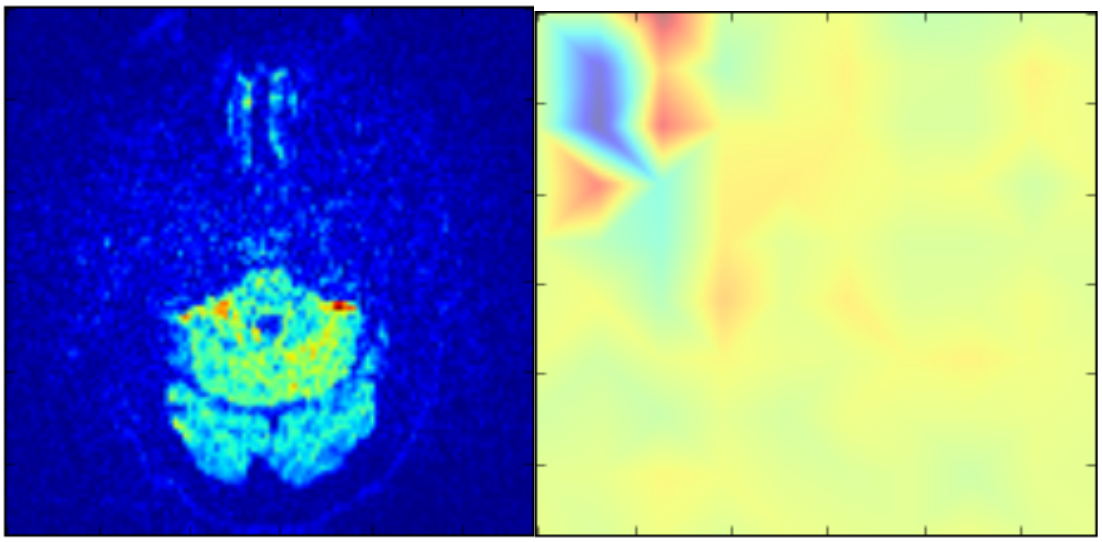

Figure 2: a) (left): Slice from original brain image at a specific time point; b) (right): Corresponding slice from tensor decomposition output

Discussion

In summary we can conclude that dimensionality reduction is a useful method for extracting meaningful features from brain imaging. Furthermore the impressive performance of these features in machine learning applications indicates that at least some subset of these features strongly correlates with the patient group. While not explored in this paper, it would be interesting to explore why LDA seemed cause a drop in classifier performance while traditional PCA did not in the tensor decomposition. In looking at Figure 2 we compare a slice of a brain image with a corresponding slice from tensor decomposition output. Because we maintain a similar color scale between the two groups, we can intuit that we have managed to select features that focus on the key areas of textural change in the image. The yellow corresponds to more diffuse, random movement of water and our slice from our decomposed core shows that we are emphasizing pixel values closely related to the region of diffuse water movement. Additionally, it would be interesting to explore the effect of various preprocessing methods to improve out comes and to systematically obscure the data to evaluate which features of the raw pixel data are being hilighted by the tensor decomposition and linear dynamical systems steps.

References

Chaudhuri, K. R., Bhidayasiri, R., & van Laar, T. (2016). Unmet needs in Parkinson's disease: New horizons in a changing landscape. *Parkinsonism & related disorders, 33*, S2-S8.

Sveinbjornsdottir, S. (2016). The clinical symptoms of Parkinson's disease. *Journal of neurochemistry, 139*(S1), 318-324.

Rabanser, S., Shchur, O., & Günnemann, S. (2017). Introduction to Tensor Decompositions and their Applications in Machine Learning. *arXiv preprint arXiv:1711.10781*.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., ... & Coggeshall, M. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet, 388*(10053), 1545-1602.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., ... & Poewe, W. (2011). The parkinson progression marker initiative (PPMI). *Progress in neurobiology, 95*(4), 629-635.

Cochrane, C. J., & Ebmeier, K. P. (2013). Diffusion tensor imaging in parkinsonian syndromes A systematic review and meta-analysis. *Neurology, 80*(9), 857-864.

Soares, J. M., Marques, P., Alves, V., & Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in neuroscience, 7*.

Chahine, L. M., & Stern, M. B. (2016). Parkinson's Disease Biomarkers: Where Are We and Where Do We Go Next?.*Movement Disorders Clinical Practice*.

Dinov, I. D., Heavner, B., Tang, M., Glusman, G., Chard, K., Darcy, M., ... & Foster, I. (2016). Predictive big data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PloS one, 11*(8), e0157077.

Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., & Zhou, J. (2017, August). Patient subtyping via time-aware lstm networks. In*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 65-74). ACM

Simuni, T., Caspell-Garcia, C., Coffey, C., Lasch, S., Tanner, C., Marek, K., & PPMI Investigators. (2016). How stable are Parkinson's disease subtypes in de novo patients: Analysis of the PPMI cohort?.*Parkinsonism & related disorders, 28*, 62-67.

Adeli, E., Wu, G., Saghafi, B., An, L., Shi, F., & Shen, D. (2017). Kernel-based Joint Feature Selection and Max-Margin Classification for Early Diagnosis of Parkinson's Disease.*Scientific reports, 7*.

Swiebocka-Wiek, J. (2016, September). Skull Stripping for MRI Images Using Morphological Operators. In*IFIP International Conference on Computer Information Systems and Industrial Management* (pp. 172-182). Springer International Publishing.

Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2016, December). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. arXiv preprint arXiv:1612.02572.

Banerjee, M., Okun, M. S., Vaillancourt, D. E., & Vemuri, B. C. (2016). A Method for Automated Classification of Parkinson's Disease Diagnosis Using an Ensemble Average Propagator Template Brain Map Estimated from Diffusion MRI. PloS one, 11(6), e0155764.

Phan, A. H., & Cichocki, A. (2010). Tensor decompositions for feature extraction and classification of high dimensional datasets. Nonlinear theory and its applications, IEICE, 1(1), 37-68.

Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A., & Moore, J. H. (2016, March). Automating biomedical data science through tree-based pipeline optimization. In European Conference on the Applications of Evolutionary Computation (pp. 123-137). Springer, Cham.

CHAPTER 3

ANALYZING LDS FEATURES AS A FUNCTION OF ORIENTATION[2]

[2] Sivakumar, R. and Quinn, S. To be submitted

Introduction

Increasingly biomedical researchers are using statistical techniques to further evaluate existing biological datasets. Part of this goal is to improve automatic classification of illnesses based on imaging, genetic, and behavioral data, but it also includes a larger goal to understand and better describe illnesses. To this end we have used an autoregressive method, linear dynamical systems (LDS) to extract features from 1196 diffusion tensor images and train a random forest classifier on these features to predict Parkinson's disease(PD) group subjects. We generate features by sampling image data at different orientation to apply to our autoregressive feature extraction protocol and successfully train a classifier with up to 94% accuracy. Additionally we show that the features produced are strongly correlated based on an average Welch's t-statistic of -0.031 across the PD group and -0.0363 across the control group.

Current biomedical research is increasingly trying to use existing medical data in new ways to aid clinicians in describing and identifying illnesses from a variety of diagnostic measures such as images, behavioral tests, and measures of blood and DNA tests. In particular it is of keen interest to researchers studying Parkinson's Disease. Previous research has looked at classifying Parkinson's disease based on brain images alone. The present study evaluates features from brain images for successfully classifying Parkinson's as well, but it also seeks to interpret the consistency of features generated by autoregressive methods.

For example, Lee and Lim (2012) evaluate gait characteristics of Parkinson's patients based on wavelet data from sensors placed on subjects' legs and feet. The researchers in this study extract several descriptive feature sets by aggregating the sensor data and use these features to successfully classify Parkinson's disease with accuracies ranging from 74 to 77 percent across their experimental protocols. This sort of methodology is valuable not just for success in

classification accuracy, but because of the generalizability of best practices for interpreting other similarly structured data for other conditions.

Likewise we have previously shown that we can successfully extract useful features for classification of Parkinson's from diffusion tensor images (DTI) using autoregressive techniques. The present study aims to show that these techniques are robust in identifying specific features that consistently correlate with each other and to the classification group.

Methods

The present study aims to examine in depth the influence of brain image orientation on features generated from autoregressive dimensionality reduction methods. We begin with a 4D pixel array of scalar maps corresponding to mean diffusion and fractional anisotropy at each voxel. If we think of movement of water as a vector these can correspond to magnitude and direction respectively. In this case these represent several 3D images taken over 65 time points at regular intervals (the number of time points is consistent across all subjects). At each time point there is a 3D image of dimensions 100, 116,116, that correspond to depth, width and height respectively. These dimensions are also consistent across subjects. To visualize how orientation of slices would affect experimental condition, please refer to Figure 3.

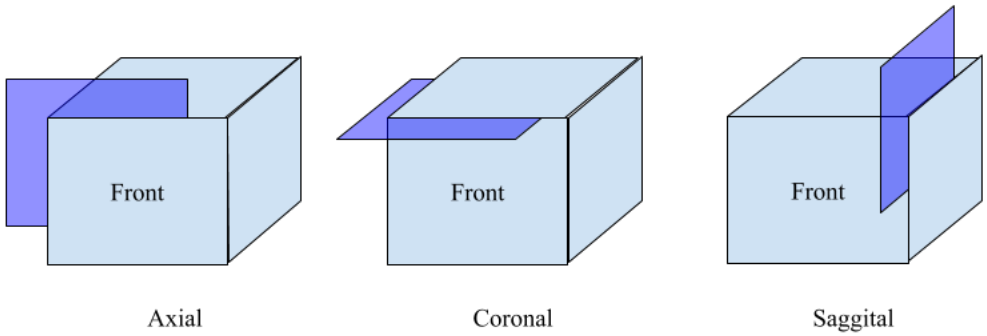

Figure 3: Illustration to show how slices are taken from the brain image for the different orientation conditions. Axial refers to vertical slices taken front to back. Coronal refers to horizontal slices. Saggital slices are vertically taken left to right.

The experiment protocol begins by iterating over corresponding slices in the 3D pixel arrays at each time point for one subject. For the example in the first case, we take coronal slices of each brain image, that is iterating over the depth axis. This would yield slices of shape 116 x 116. We would grab one at each time point and flatten to a vcctor. Stacking these would yield a 2D matrix of which one dimension is number of time points. From this we perform singular value decomposition to reduce the dimensionality of the matrix. We further reduce the time dimension by calculating the transition model (Hyndman, et al., 2007). We then flatten to a vector and repeat for all slices along the dimension. Each additional vector is appended to form one final feature vector for each subject. These are then used as training features for a random forest classifier. We have chosen random forest based on a genetic algorithm based pipeline evaluator. Based on several generations of classifier pipelines, the algorithm found random forest classifiers to be optimal based on accuracy and f-measure.

Results

Our classification results as do not differ significantly based on changes in orientation of slice for saggital and axial slices, while coronal slices performed slightly worse. As we can see from Tables 3 through 5, we consistently perform well on features extracted from the images without additional post-processing, presumably because we lose information as perform secondary dimensionality reduction on the features. To verify how similar the features are, we perform a subject to subject Welch's T-test on the features between each possible pairing of the orientation groups. We average the results across the subject groups. We are thus able to see that the features are statistically similar regardless of which orientation we iterate over.

We do note however a marked decrease in performance in the coronal slice condition relative to the other two. It is possible this effect may be due to the upward progression of Parkinson's. As we take slices and reduce their dimensionality, we inevitably lose more key information on key slices that are predominated with affected tissue. These may conversely rise to the top in vertical slices which equally cover neocortical regions less likely to be affected by Parkinson's and lower regions more likely to be affected.

Table 3: Cross-validated(n=5) classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied. (Orientation 1, coronal slices)

Dimensionality Reduction Method	F-measure	Accuracy
-	0.87±0.007	0.77±0.001
PCA	0.83±0.009	0.72±0.014
LDA	0.83±0.024	0.73±0.034
Kernel PCA	0.84±0.009	0.72±0.013

Table 4: Cross-validated(n=5) classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied. (Orientation 2, sagittal slices)

Dimensionality Reduction Method	F-measure	Accuracy
-	0.98±0.021	0.98±0.022
PCA	0.89±0.019	0.90±0.022
LDA	0.85±0.015	0.85±0.017
Kernel PCA	0.91±0.040	0.91±0.039

Table 5: Cross-validated(n=5) classification accuracy of features generated from linear dynamical systems after various additional dimensionality reduction techniques are applied. (Orientation 3, axial slices)

Dimensionality Reduction Method	F-measure	Accuracy
-	0.98±0.021	0.98±0.022
PCA	0.91±0.023	0.91±0.023
LDA	0.82±0.019	0.84±0.020
Kernel PCA	0.91±0.028	0.91±0.028

Table 6: Mean Welch's T-statistics for Parkinson's and control groups when comparing features from different slice orientations.

Orientation	**Parkinson's Disease**	**Control**
Coronal v Saggital	0.0072061	0.025465
Saggital v Axial	-0.057516	-0.083675
Coronal v Axial	-0.045403	-0.050656

Table 7: Resulting p-value corresponding to mean t-statistic comparing orientation as function of patient condition

Orientations	**Parkinson's Disease**	**Control**
Coronal v Saggital	0.41136	0.41927
Saggital v Axial	0.41136	0.41927
Coronal v Axial	0.46706	0.50251

Discussion

In summary we show that features extracted through autoregressive methods can be consistently used to evaluate brain images from Parkinson's and control subjects. The success across all orientation types suggests that in all cases features being extracted are strongly correlated to patient group. Additionally, the Welch's T-test shows that this methodology produces similar t-statistic for vertically sliced orientations. This is encouraging with regards to the generalizability of this approach to feature extraction. These promising results suggest that autoregressive methods may be appropriate means of feature extraction in other imaging modalities as well as in other types of biomedical data. These could also be a useful tool for consistently, and efficiently quantifying images and other data from patients.

On the other hand, the coronal slices (taken horizontally) performed significantly worse, even though statistically, the features this orientation generated aren't significantly different from those produced from the other two orientations. Table 6 and 7 illustrate this when we look at the mean

T-statistic (assuming unequal variance) between the possible pairwise comparisons. We fail to reject the null hypothesis, as we see very high p-values and low t-statistic.

References

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2007). Diffusion tensor imaging of the brain. Neurotherapeutics, 4(3), 316-329.

ROSENBROCK, H. H. (1974). Structural properties of linear dynamical systems. International Journal of Control, 20(2), 191-202.

Lee, S. H., & Lim, J. S. (2012). Parkinson's disease classification using gait characteristics and wavelet-based feature extraction. Expert Systems with Applications, 39(8), 7338-7344.

Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A., & Moore, J. H. (2016, March). Automating biomedical data science through tree-based pipeline optimization. In European Conference on the Applications of Evolutionary Computation (pp. 123-137). Springer, Cham.

Hyndman, M., Jepson, A. D., & Fleet, D. J. (2007, September). Higher-order Autoregressive Models for Dynamic Textures. In BMVC (pp. 1-10).

CHAPTER 4

CONCLUSION

In conclusion it is clear there is clear value in using pixel based features of brain imaging to better understand and quantify diseases such as Parkinson's. In doing so, not only can we more consistently detect biomarkers and correlate them to the onset of outward manifestations of the disease. This has implications for other diseases, not just those of neurological origin.

Features obtained from extracting a core tensor via Tucker decomposition were able to train a random forest classifier to successfully classify Parkinson's brain images at 94% accuracy. These results are further validated with a strong f-measure score of 94% as well.

When extracting features using linear dynamical systems, we are once again able to achieve very strong results. We are able to train a random forest classifier to 82% accuracy at an f-measure score of 90%. These results are further boosted by further reducing the extracted features vectors using kernel PCA. These reduced features are able to train a random forest classifier to predict Parkinson's from brain images with 92% accuracy and f-measure of 93% By also attempting to extract features through varying orientations, we are able to increase baseline accuracy to 94% with rebalanced classes. We also show that these features are effectively similar in mean and variance using Welch's t-test on a subject to subject basis between orientation groups. When reviewing the mean statistics in the Parkinson's and control groups, we see that we can not reject the null hypothesis that the mean of the features differ according to orientation.

REFERENCES

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2007). Diffusion tensor imaging of the brain. Neurotherapeutics, 4(3), 316-329.

Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., & Del Tredici, K. (2004). Stages in the development of Parkinson's disease-related pathology. Cell and tissue research, 318(1), 121-134.

Del Tredici, K., & Braak, H. (2016). Sporadic Parkinson's disease: development and distribution of α-synuclein pathology. Neuropathology and applied neurobiology, 42(1), 33-50.

Chaudhuri, K. R., Bhidayasiri, R., & van Laar, T. (2016). Unmet needs in Parkinson's disease: New horizons in a changing landscape. *Parkinsonism & related disorders, 33*, S2-S8.

Sveinbjornsdottir, S. (2016). The clinical symptoms of Parkinson's disease. *Journal of neurochemistry, 139*(S1), 318-324.

Rabanser, S., Shchur, O., & Günnemann, S. (2017). Introduction to Tensor Decompositions and their Applications in Machine Learning. *arXiv preprint arXiv:1711.10781*.

Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., ... & Coggeshall, M. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet, 388*(10053), 1545-1602.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., ... & Poewe, W. (2011). The parkinson progression marker initiative (PPMI). *Progress in neurobiology, 95*(4), 629-635.

Cochrane, C. J., & Ebmeier, K. P. (2013). Diffusion tensor imaging in parkinsonian syndromes A systematic review and meta-analysis. *Neurology, 80*(9), 857-864.

Soares, J. M., Marques, P., Alves, V., & Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Frontiers in neuroscience, 7*.

Chahine, L. M., & Stern, M. B. (2016). Parkinson's Disease Biomarkers: Where Are We and Where Do We Go Next?.*Movement Disorders Clinical Practice*.

Dinov, I. D., Heavner, B., Tang, M., Glusman, G., Chard, K., Darcy, M., ... & Foster, I. (2016). Predictive big data analytics: a study of Parkinson's disease using large, complex, heterogeneous, incongruent, multi-source and incomplete observations. *PloS one, 11*(8), e0157077.

Baytas, I. M., Xiao, C., Zhang, X., Wang, F., Jain, A. K., & Zhou, J. (2017, August). Patient subtyping via time-aware lstm networks. In*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 65-74). ACM

Simuni, T., Caspell-Garcia, C., Coffey, C., Lasch, S., Tanner, C., Marek, K., & PPMI Investigators. (2016). How stable are Parkinson's disease subtypes in de novo patients: Analysis of the PPMI cohort?.*Parkinsonism & related disorders, 28*, 62-67.

Adeli, E., Wu, G., Saghafi, B., An, L., Shi, F., & Shen, D. (2017). Kernel-based Joint Feature Selection and Max-Margin Classification for Early Diagnosis of Parkinson's Disease.*Scientific reports, 7*.

Swiebocka-Wiek, J. (2016, September). Skull Stripping for MRI Images Using Morphological Operators. In*IFIP International Conference on Computer Information Systems and Industrial Management* (pp. 172-182). Springer International Publishing.

Cole, J. H., Poudel, R. P., Tsagkrasoulis, D., Caan, M. W., Steves, C., Spector, T. D., & Montana, G. (2016, December). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *arXiv preprint arXiv:1612.02572.*

Banerjee, M., Okun, M. S., Vaillancourt, D. E., & Vemuri, B. C. (2016). A Method for Automated Classification of Parkinson's Disease Diagnosis Using an Ensemble Average Propagator Template Brain Map Estimated from Diffusion MRI. *PloS one*, 11(6), e0155764.

Gui, X.., Chuansheng, C.., Zhong-Lin, L.., & Qi, D.. (2010). Brain Imaging Techniques and Their Applications in Decision-Making Research. Xin li xue bao. Acta psychologica Sinica, 42(1), 120.

Dale, A. M., & Halgren, E. (2001). Spatiotemporal mapping of brain activity by integration of multiple imaging modalities. Current opinion in neurobiology, 11(2), 202-208.

Martínez-Murcia, F. J., Górriz, J. M., Ramírez, J., Illán, I. A., & Puntonet, C. G. (2013, June). Texture features based detection of Parkinson's disease on DaTSCAN images.

In International Work-Conference on the Interplay Between Natural and Artificial Computation (pp. 266-277). Springer, Berlin, Heidelberg.

Lewine, J. D., Davis, J. T., Bigler, E. D., Thoma, R., Hill, D., Funke, M., ... & Orrison, W. W. (2007). Objective documentation of traumatic brain injury subsequent to mild head trauma: multimodal brain imaging with MEG, SPECT, and MRI. The Journal of head trauma rehabilitation, 22(3), 141-155.

Phan, A. H., & Cichocki, A. (2010). Tensor decompositions for feature extraction and classification of high dimensional datasets. Nonlinear theory and its applications, IEICE, 1(1), 37-68.

Olson, R. S., Urbanowicz, R. J., Andrews, P. C., Lavender, N. A., & Moore, J. H. (2016, March). Automating biomedical data science through tree-based pipeline optimization. In European Conference on the Applications of Evolutionary Computation (pp. 123-137). Springer, Cham.

Kossaifi, J., Panagakis, Y., Anandkumar, A., & Pantic, M. (2019). Tensorly: Tensor learning in python. The Journal of Machine Learning Research, 20(1), 925-930.

Ruiters, R., & Klein, R. (2009, June). BTF compression via sparse tensor decomposition. In Computer Graphics Forum (Vol. 28, No. 4, pp. 1181-1188). Oxford, UK: Blackwell Publishing Ltd.

Hyndman, M., Jepson, A. D., & Fleet, D. J. (2007, September). Higher-order Autoregressive Models for Dynamic Textures. In BMVC (pp. 1-10).