Journal of Neuroscience Methods xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

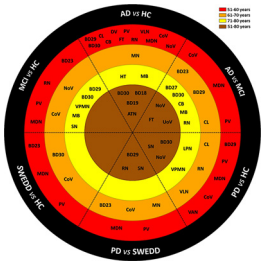
journal homepage: www.elsevier.com/locate/jneumeth

Determination of Imaging Biomarkers to Decipher Disease Trajectories and Differential Diagnosis of Neurodegenerative Diseases (DIsease TreND)

Gurpreet Singh[a,b,1], Lakshminarayanan Samavedham[a,c,*], Erle Chuen-Hian Lim[d], , the Alzheimer's Disease Neuroimaging Initiative[2], the Parkinson Progression Marker Initiative[2]

[a] Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore
[b] Department of Radiology, Dalio Institute of Cardiovascular Imaging, NewYork-Presbyterian Hospital and the Weill Cornell Medicine, New York, United States
[c] Residential College 4, 8 College Avenue West, #02-16W, Education Resource Centre, Singapore 138608, Singapore
[d] Department of Neurology, National University Health System, National University of Singapore, Singapore

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Neurodegenerative disease diagnosis
Computer-assisted diagnosis
Differential diagnosis
Imaging biomarkers
Machine learning
Unsupervised learning

ABSTRACT

Background: Understanding disease progression of neurodegenerative diseases (NDs) is important for better prognosis and decisions on the appropriate course of treatment to slow down the disease progression.
New Method: We present here an innovative machine learning framework capable of (1) indicating the trajectory of disease progression by identifying relevant imaging biomarkers and (2) automated disease diagnosis. Self-Organizing Maps (SOM) have been used for data dimensionality reduction and to reveal potentially useful disease-specific biomarkers, regions of interest (ROIs). These ROIs have been used for automated disease diagnosis using Least Square Support Vector Machines (LS-SVM) and to delineate disease progression.
Results: A multi-site, multi-scanner dataset containing 1316 MRIs was obtained from ADNI[3] and PPMI. Identified biomarkers have been used to decipher (1) trajectory of disease progression and (2) identify clinically relevant ROIs. Furthermore, we have obtained a classification accuracy of 94.29 ± 0.08% and 95.37 ± 0.02% for

Abbreviations: AN, anterior nucleus of thalamus; BD18, Brodmann area 18; BD19, Brodmann area 19; BD23, Brodmann area 23; BD27, Brodmann area 27; BD29, Brodmann area 29; BD30, Brodmann area 30; BD33, Brodmann area 33; CB, caudate body; CL, cerebellar lingual; CoV, culmen of vermis; DV, Declive; FT, fastigium; HT, hypothalamus; LPN, lateral posterior nucleus; MB, mammillary Body; MDN, medial dorsal nucleus; MN, midline nucleus; NoV, nodule of vermis; PV, pulvinar; RN, red nucleus; SN, substantia nigra; UoV, uvula of vermis; VAN, ventral anterior nucleus; VLN, ventral lateral nucleus; VPMN, ventral posterior medial nucleus; AD, Alzheimer disease; PD, Parkinson disease; MCI, mild cognitive impairment; SWEDD, patients with scans without evidence of dopaminergic deficit; HC, healthy controls

* Corresponding author at: Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore.
E-mail addresses: gus2011@med.cornell.edu (G. Singh), laksh@nus.edu.sg (L. Samavedham), mdcelch@nus.edu.sg (E.C.-H. Lim).

[1] Presently working as Cognitive Software Engineer at Dalio Institute of Cardiovascular Imaging, NewYork-Presbyterian Hospital and the Weill Cornell Medicine, New York.

[2] Data collection and sharing for this study was funded by the Alzheimer's disease Neuroimaging Initiative (http://adni.loni.usc.edu/) and the Parkinson's Progression Markers Initiative (http://www.ppmi-info.org/) primarily supported by National Institute on Aging, the National Institute of Biomedical Imaging & Bio-engineering and by the Michael J. Fox Foundation for Parkinson's Research respectively. The investigators within the ADNI and PPMI only provided the data but they did not contribute towards analysis or writing of this report.

[3] **ADNI**: Alzheimer's disease neuroimaging initiative and **PPMI:** Parkinson's progression markersinitiative.

https://doi.org/10.1016/j.jneumeth.2018.05.009
Received 15 June 2017; Received in revised form 30 January 2018; Accepted 14 May 2018

distinguishing AD and PD from HC subjects respectively.

Comparison with other existing methods: The goal of this study was fundamentally different from other machine learning based studies for automated disease diagnosis. We aimed to develop a method that has two-fold benefits (1) It can be used to understand pathology of neurodegenerative diseases and (2) It also achieves automated disease diagnosis.

Conclusions: In the absence of established disease biomarkers, clinical diagnosis is heavily prone to misdiagnosis. Being clinically relevant and readily adaptable in the current clinical settings, the developed framework could be a stepping stone to make machine learning based Clinical Decision Support System (CDSS) for neurodegenerative disease diagnosis a reality.

1. Introduction

The aging population is a major global demographical trend with great social, political and economic consequences. In the next four decades, the share of people aged 60 and above is expected to rise to 22% of the total population, a jump from 841 million to 2.1 billion people (United Nations, 2013). This phenomenon is closely linked to an increased prevalence of dementia. Alzheimer disease (AD) is the leading cause of dementia in the aged population, whereas Parkinson disease (PD) is the fifth most common form of dementia and one of the leading cause for movement related disorder. Both of these NDs are expected to double in numbers over the next two decades (Dorsey et al., 2007; Prince et al., 2013). Currently, there is no cure for NDs.

Over the last decade, there has been a growing interest in machine learning based approaches amongst the neuroimaging community. Using machine learning, classifier models are built by extracting information from imaging modalities such as functional Magnetic Resonance Images (fMRI) (Chen et al., 2011; Kenny et al., 2012; Li et al., 2014), Positron Emission Topography (PET) (Bailly et al., 2015; Nagano-Saito et al., 2004), Diffusion Tensor Imaging (DTI) (Graña et al., 2011), structural Magnetic Resonance Imaging (MRI) (Salvatore et al., 2014; Singh and Lakshminarayanan, 2015) and others. Of these modalities, structural MRI has emerged to be a useful indicator of disease progression, is non-invasive, widely available, and have the potential to assist with clinical diagnosis and monitoring the progression of disease in most, if not all, of the NDs (Jack et al., 2010; Kassubek and Müller, 2016).

Machine learning based methods are typically suited for handling high dimensional datasets, like neuroimaging modalities. Machine learning tools and statistical methods such as Principal Component Analysis (PCA) (Salvatore et al., 2014), Independent Component Analysis (ICA) (Yang et al., 2014) and Pearson's correlation (Graña et al., 2011) have been used for feature extraction. For subject classification, these features may be presented to classification algorithms such as Support Vector Machines (SVM). By combining neuroimaging modalities with machine learning algorithms, a new era of intelligent non-knowledge based clinical decision support systems (CDSS) is emerging.

However, studies for automated disease diagnosis are rarely designed to provide clinically interpretable insights. Determination of reliable biomarker(s), that could be used to identify and track the progression of neurodegenerative disease(s), could be a better target for devising clinically applicable machine learning tools for neuroimaging based disease diagnosis as they could also help in understanding disease pathology. Our devised methodology has two-fold benefits wherein a unique combination of unsupervised and supervised learning algorithms has been used

1 To identify ROIs that correspond to regions known to be affected by a disease, imaging biomarker, without the availability of a priori information and
2 To obtain high classification accuracy on the largest dataset reported in the literature for machine learning based approaches for diagnosis of NDs

On MRIs, changes in the brain appear as the change in RGB color channel values. However, the location and amount of this change are localized and small as compared to the overall information present on MRIs. Thus, the need for a technique that can not only do feature extraction but also preserve the topological representation. For feature extraction, we have made use of self-organizing maps (SOM). SOM is an unsupervised learning based machine learning algorithm for non-linear dimensionality reduction well suited for large multi-dimensional datasets (Kohonen, 1998), like brain MRIs. It produces a representation of patterns in the input data by mapping it to an output space such that the nodes in the output map represent physical spatial patterns that can directly be reversibly mapped to input space unlike traditional methods such as PCA. Further, it is inherently suited for generating a low-dimensional representation of high-dimensional datasets while still preserving the inherent non-linearity and underlying distribution of data. This introduction of nonlinearity not only provides a generalized pattern of the data but also improves the topographical optimization and prevents overfitting to the data, making it less prone to noise in the data.

Many brain regions affected by NDs appear as signal intensity abnormalities on MRIs. The presence of signal intensity abnormalities in a specific pattern or at a specific location is indicative of a kind of disease. However, Machine learning studies have been conducted in an isolated manner focused only on one of the two major neurodegenerative diseases at a time *i.e.* either AD or PD. There is an increasing evidence that NDs exhibit overlapping clinical symptoms during early stages of NDs, such as AD with subjects suffering from Mild Cognitive Impairment (MCI) (Walker et al., 2015; Xie et al., 2014). The lack of established biomarkers and similar clinical presentation makes it difficult to diagnose NDs at early stages of the disease in a clinical setting. Studying these diseases together is important to understand their epidemiology and also to assess the generalizability of a methodology. Selection of clinically relevant ROIs from these images, that can help to differentiate between two diseased subjects or a diseased subject from a healthy control, could be the key towards designing more practical CDSS. Availability of such a CDSS capable of disease prognostication by comparing the brain areas affected by disease trajectory bio-markers could greatly improve the clinical efficiency by assisting in disease prognosis. Thus, from practical perspective, it is imperative to have a CDSS capable of

1 Detecting changes in the brain tissue
2 Understanding common disease progression pathways and
3 Classifying patients in terms of predefined patient classes

To achieve this, we have used machine learning based approach to identify relevant regions for differential diagnosis of the AD, PD, Mild Cognitive Impairment (MCI), Scans Without Evidence of Dopaminergic Deficit (SWEDD) and Healthy Control (HC) subjects. We have developed an innovative framework wherein unsupervised learning based Self-Organising Maps (SOM) has been used to determine unknown, but the potentially useful structure in the imaging data, Regions of Interest