International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 3 Issue: 10
ISSN: 2454-4248
181 – 187

Progression Modeling of Cognitive Disease Using Temporal Data Mining: Research Landscape, Gaps and Solution Design

Dr. Veenu Mangat
UIET, Panjab University
Chandigarh, India
veenumangat@yahoo.com

***Abstract*—**Dementia is a cognitive disorder whose diagnosis and progression monitoring is very difficult due to a very slow onset and progression. It is difficult to detect whether cognitive decline is due to ageing process or due to some form of dementia as MRI scans of the brain cannot reliably differentiate between ageing related volume loss and pathological changes. Laboratory tests on blood or CSF samples have also not proved very useful. Alzheimer's disease (AD) is recognized as the most common cause of dementia.

Development of sensitive and reliable tool for evaluation in terms of early diagnosis and progression monitoring of AD is required. Since there is an absence of specific markers for predicting AD progression, there is a need to learn more about specific attributes and their temporal relationships that lead to this disease and determine progression from mild cognitive impairment to full blown AD. Various stages of disease and transitions from one stage to the have be modelled based on longitudinal patient data.

This paper provides a critical review of the methods to understand disease progression modelling and determine factors leading to progression of AD from initial to final stages. Then the design of a machine learning based solution is proposed to handle the gaps in current research.

Keywords- *Disease Progression Modeling; Temporal Pattern mining; Predictive Accurac; Normal Cognition; Mild Cognitive Impairment; Alzheimer's Disease*

*****

I. INTRODUCTION

Dementia is a term to describe a set of neuropsychiatric symptoms that can include memory loss and difficulties with thinking, problem-solving or language. These symptoms occur when the brain is damaged by certain diseases, including Alzheimer's disease (AD). Chronic neurodegenerative disorders such as AD are also well known risk factors for depression. AD is the most common cause of dementia accounting for nearly 60 to 80 percent of cases. The early symptoms associated with AD are difficulty remembering recent conversations, names or events, apathy, depression, etc. Later symptoms include impaired communication, poor judgment, disorientation, confusion, behavior changes and difficulty speaking, swallowing and walking. AD is now considered a slowly progressive brain disease that begins well before symptoms emerge. Hence understanding its progression is important.

The modeling of a disease using computational methods is known as disease progression modeling. Understanding how a disease progresses is fundamental to early diagnosis, preventive medicine and personalized care. Understanding the temporal progression of diseases leads to improvement in disease prognosis, development of drug therapy and design of clinical trials.

Current data science research is advancing conversion of medical data to aid in detection, diagnosis and treatment of diseases. The immense potential of machine learning and data mining should be harnessed to obtain models of meaningful and insightful data pertaining to chronic, degenerative and life threatening diseases from the vast amounts of high dimensional data being recorded through Electronic Medical Records. Electronic Medical Records typically contain diagnosis, medication and clinical notes. Longitudinal patient health information in the form of demographics, encounter records, medications, vital signs, laboratory data, immunizations, etc. is recorded in the EMRs.

The purpose of this paper is to review existing solutions and propose the use of a machine learning based solution that can identify factors/biomarkers and infer temporal progression of diseases based on recorded observations. It is proposed to implement longitudinal feature learning and analyze temporal patterns of clinical factors and/or biomarkers in disease progression.

One of the main issues in deriving a temporal model for diseases based on recorded evidence is the incompleteness and sparsity of observations. Medical records often cover only part of the disease progression path. Also, not all variables of interest are recorded for all patients. So the proposed solution should include sparse learning techniques. Also medical data is often large scale and high dimensional in nature. The model can employ some multitask learning formulations for disease progression.

Secondly, observations are mostly irregular with respect to time for different patients. The time granularity varies across patient records. Hence, some method of imposing temporal smoothness on the data without losing any crucial data needs to be devised.

The third issue is the non uniformity or heterogeneity of patient disease progression rates. The observations are discrete with non equal intervals, which pose an additional

 ISSN: 2454-4248
181 – 187

challenge in mining a smooth disease progression trajectory. Also, identifying multiple covariates, which may be hidden, is non trivial. Additionally, data is large scale and may contain noise. The machine learning approach should have a mechanism to handle all these issues.

It is proposed to apply the approach to study of AD since it is the most common form of the cognitive disorder: dementia. AD is a complex disease associated with synapse loss and neurodegeneration leading to memory impairment and other cognitive problems. There is currently no known treatment that slows the progression of this disorder. According to the 2010 World Alzheimer report, there are an estimated 35.6 million people worldwide living with dementia at a total cost of more than US$600 billion in 2010, and the incidence of AD throughout the world is expected to double in the next 20 years. There is a pressing need to find biomarkers to both predict future clinical decline and for use as outcome measures in clinical trials of disease-modifying agents.

II. Review of Related Work

A. Review of Work in Disease Progression Modeling

A multistage Hidden Markov Model for estimating transition rates and probabilities of misclassification was discussed in [1] and applied to an aneurysm screening study. A model based on meta-analysis to describe the longitudinal changes in parameters of patients with mild to moderate Alzheimer's disease was proposed in [2]. A dynamic Bayesian network based technique was used to model the progression of Coronary Atherosclerosis in [3]. A mechanism for modeling progression of diabetes mellitus by tracking interaction between key factors was developed in [4]. A group of models to describe degenerative diseases as a function of diseases process and effects of treatment given, was proposed in [5]. All these models were specifically targeted at a particular disease and included significant domain knowledge of the main factors of the disease under study and its progression.

The Hidden Markov Model was applied to study stages of Alzheimer's disease in [6]. This was a discrete time model and was unable to handle irregular time granularities of patient visits. Physiological, clinical and treatment variables were collected at fixed time intervals in an ICU and then clustered using hierarchical clustering methods in order to recognize patient states [7]. Individual patient states were visualized over time.

A novel statistical kernel model designed to learn the complex glycan interactions and predict the differences in AIDS disease progression using the structural 3D glycan profile, was proposed in [8]. It involves the design of semi-parameterized, and support vector assisted hierarchical mixture model, which is able to effectively capture the information of non-local interactions with strong resistance to vanishing gradient and high-dimensionality problems. The proposed framework successfully classified the changes to glycosylation profiles and segregated HIV disease groups. These results show the utility of new bioinformatics and machine learning tools in providing useful biological understanding of glycosylation patterns during AIDS disease progression.

A fused group lasso formulation for modeling disease states with known biomarkers was proposed in [9]. This work was refined in [10] by using multitask learning formulation instead of single task learning. A technique based on joint time sequence segmentation was used to study stages of a disease by considering each patient EMR as a time sequence [11]. An unsupervised method for evidence based modeling of disease based on longitudinal clinical findings of a cohort of patients was proposed in [12]. The proposed model used a Markov jump process to model transition of stages of a disease wherein the temporal patterns of comorbidities determine the transitions. The presence of a comorbidity is determined by using a bipartite Bayesian network. The method only considers patients having the disease under study in its training phase.

A data driven phenotyping framework was proposed in [13]. Each medical concept was viewed as a combination of several observed medical features. Two approaches namely Individual Basis Approach and Shared Basis Approach, were developed assuming phenotypes are different for each patient in the former case and phenotypes are common to patient population in the latter case respectively. A scalable and efficient Block Coordinate Descent optimization method was used to tackle the two formulations. The multitask learning paradigm in the form of inductive transfer learning was used for Shared Basis Approach. Different approaches and challenges for prediction modeling from patient data were discussed in [14]. These methods construct feature vectors with summary statistics and then input the feature vectors so derived into models for classification, clustering, regression, etc. But these models do not work well with sparse and noisy data.

Authors of [15] presented a framework for classification of multivariate time series analysis, which is composed of three phases. The first phase applies a temporal abstraction process to transform a series of raw time-stamped data points into a series of symbolic time intervals (based on either unsupervised or supervised temporal abstraction). The next phase mines these time intervals to discover frequent temporal-interval relation patterns. Finally these temporal patterns are used as features to induce a classifier.

B. Review of Related Work in Alzheimer's Disease Progression Modeling

One approach toward an understanding of Alzheimer's Disease is the formulation of a disease model [16-18]. Jack et al [17][18] presented a hypothetical model for biomarker dynamics in AD pathogenesis. The model begins with the abnormal deposition of Aβ fibrils, and sometime later, neuronal damage begins to occur. As neuronal degeneration progresses, atrophy in certain areas typical of AD becomes detectable by magnetic resonance imaging (MRI). The model provided by Jack et al [17] is highly relevant to many studies of the ADNI cohort which provide empirical evidence to support it. It is now generally accepted that the initial AD pathology develops in situ while the patient is cognitively normal, known as the 'preclinical stage' [19]. At some point in time, sufficient brain damage accumulates to

result in cognitive symptoms and impairment. Originally defined in 1999, this has been classified in a number of ways, including as predementia AD or as mild cognitive impairment (MCI), or as the prodromal state AD [20].

Although considerable work had been done to develop quantitative measurements for clinical trials of symptomatic treatments, it was recognized that changes in cognition did not necessarily signify disease modification. Therefore, investigators from academia and the pharmaceutical industry became interested in how disease modification of AD could be detected using a variety of biomarkers, including brain MRI scanning, and blood and CSF analytes. This led to a decision by the National Institute on Aging (NIA) to fund the ADNI. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is an ongoing, longitudinal, multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD [21]. Now in its third phase, ADNI2 is studying rate of change of cognition, function, brain structure and biomarkers in 150 elderly controls, 450 subjects with mild cognitive impairment, 150 with mild to moderate AD and a new group of 100 people with significant memory concern cohort.

Trojanowski et al [22] reviewed progress by the Biomarker Core of ADNI in developing profiles of CSF or plasma biomarkers that would act as a "signature" of mild AD or predict future MCI to AD conversion. Moreover, the review described studies in support of a temporal sequence of changes in individual biomarkers that reflected proposed trajectories of Aβ deposition and the formation of neurofibrillary tangles in AD progression [17].

Authors of [23] distinguished between diagnostic biomarkers and risk biomarkers, such as the APOE ε4 allele and plasma total homocysteine levels, suggesting that although these were not sufficiently sensitive for diagnostic purposes, yet were indicative of increased risk for AD and were predictive of disease progression.

The accomplishments of the Clinical Core of ADNI were reviewed by Aisen et al [24], who reported that the Core had successfully recruited a cohort of >800 subjects, characterizing them both clinically and cognitively at baseline and following them longitudinally over the course of the study.

Correctly distinguishing patients with AD pathology is critical, especially considering the overlap that exists between various late-life neurodegenerative pathologies. For example, the Lewy bodies that characterize Parkinson's disease are found in >50% of patients with AD, in addition to neuritic plaques and tangles. Therefore, there is a real need for biomarkers that reliably distinguish between different types of dementias [25].

The idea that rates of change of brain atrophy and hypometabolism are not uniform but vary by disease stage is supported by several studies [26][27]. Moreover, the rates of change differ among the various groups.

A machine learning algorithm proposed by Hinrichs et al [28] integrates a spatial discrimination step to identify AD-related patterns in different brain regions, rather than assessing these relationships at the pre- or postprocessing steps. The development of a panoply of multimodal classifiers that leverage information from imaging, biological and neuropsychological sources has been a major focus of ADNI papers. Likewise, the selection of features that are most 'AD-like' across multiple modalities is a critical step in constructing an accurate classifier and new approaches to this step have been reported. Hinrichs et al [29] developed a method based on the Multi-Kernel Learning framework to produce a classifier that, in addition to classifying control and AD patients, also produced a Multi-Modality Disease Marker (MMDM) that could be used for the prediction of MCI to AD conversion. This method consistently outperformed a similarly trained SVM using the ADNI data set.

An alternative method for AD classification that uses a non-negative matrix factorization for feature selection in combination with SVMs with bounds of confidence for classification was reported by Padilla et al [30]. The authors found that this method was an accurate tool for classifying AD patients from a combination of SPECT and PET data.

Zhang et al [31] reported the first work to combine not only imaging but also biological data in the form of levels of CSF biomarkers into multi-modal classifier. They used a linear SVM with an intrinsic feature selection mechanism to rank top features of 93 ROIs and CSF biomarkers were added directly as features. This method achieved high classification accuracy.

The next step in utilizing these classifiers is to determine their effectiveness in the prediction of future cognitive decline in addition to classification problems. Combining MR, FDG-PET and CSF data is the focus of [32] wherein a method, multi-modal multi-task (M3T), that uses this disparate data to estimate both continuous variables, such as scores on neuropsychological tests (MMSE, ADAS-cog), by regression and a categorical variable, has been proposed. M3T combines a multi-task feature selection with a multi-modal SVM that fuses selected features for regression and classification. M3T was found to be more effective as comparable to other reported prediction methods [33].

To account for the fact that brain structures in imaging data are interconnected, Wang et al [34] proposed the Sparse Multi-task Regression and Feature Selection (SMART) method that jointly analyzed all imaging and clinical data using a single regression model with sparse multi-task learning, and found that this method was an improvement on multi-variate regression when used to predict decline in AVLT scores. Mattila et al [35] and Soininen [36] created a diagnostic decision support system by representing cognitive, imaging, biological and genetic data in a graphical form termed a Disease State Fingerprint (DSF), as well as statistically distilling a score, the Disease State Index (DSI), that reflects the likelihood of a patient having AD. Meda et al [37] presented a method for multivariate analysis of GWAS data based on the premise that genetic determinants are not randomly distributed throughout the genome, but tend to cluster in specific biological processes related to AD. Their method used a parallel ICA and a hypothesis-free, data-driven statistical technique to simultaneously examine multiple modalities. Wang et al [38] developed a novel method, Group-Sparse Multi-task Regression and Feature selection (G-SMuRFS) that is built on multivariate regression analysis with a new form of regularization. Application of the method using the ADNI

 ISSN: 2454-4248
181 – 187

data-set demonstrated its ability to predict continuous responses of brain imaging measures and to select relevant SNPs in a more efficient manner than conventional multivariate linear regression.

A variety of other approaches have been used to diagnose MCI and AD, some based on single measures, others on composite scores of a single modality, and still others on a combination of factors from different modalities. Although most classifiers used baseline measurements, there is some evidence to suggest that longitudinal data may provide even more accurate diagnoses, but it remains to be seen whether this approach is more generally applicable to other modalities. Currently, the best classifiers are able to discriminate between control and AD subjects with accuracies in the mid-90% range, but have considerably lower accuracies when discriminating between control and MCI subjects or between MCI-nc and MCI-c subjects, although data for the latter diagnoses, are far less reported.

In [39], longitudinal ADAS-cog data from ADNI was used to construct a model that included baseline severity, APOE status, age and gender identified as covariates to predict a curvilinear rate of disease progression. Samtani et al [40] also used longitudinal changes in ADAS-cog scores and developed a non-linear mixed effects model for disease progression in AD. Llano et al [41] used a new Random Forests tree-based multivariate model of ADAS-cog in which the subscores had been weighted according to their contribution to patient discrimination. This model, ADAS.Tree, predicted conversion of MCI to AD more accurately than baseline MMSE or ADAS-cog and, in addition, was a better predictor of conversion.

There are various ongoing projects [42] based on ADNI in institutes like Johns Hopkins School of Public Health, Imperial College, Stanford, etc. Eg: King's College London is focusing on improved estimation of rate of decline and subsequent detection of biomarkers of AD progression.

Authors of [43] have extracted geometric features such as minor axis, Euler number and solidity from Alzheimer brain magnetic resonance images in order to demarcate the Alzheimer subjects from the control normals. They have analyzed the atrophy of Corpus Callosum (CC) in using anisotropic diffusion filtering and modified distance regularized level set method. Anisotropic diffusion filtering is used as preprocessing to obtain the edge map. The modified distance regularized level set method is employed to segment CC using this edge map. Geometric features are extracted from the segmented CC and are analyzed. Results show that anisotropic diffusion filtering is able to extract the edge map with high contrast and continuous boundaries. Modified distance regularized level set method could perform the segmentation of CC in both normal and Alzheimer images.

Minkowski functionals (MFs) based brain to ventricle index is proposed in [44] to quantify the structural changes in Alzheimer's MR brain images for severity detection. The original images are skull stripped using morphology based method. The ventricles are segmented using localized region-based active contour method. Results show that the MF2D brain to ventricle index provide better discrimination of normal and abnormal subjects and its correlation with MMSE is observed to be high for normal and very high for Alzheimer subjects. This index can be used to quantify the progression in neurodegenerative disorder such as AD.

The shape changes of CC have been analysed using shape based Laplace Beltrami (LB) eigen value features and machine learning techniques in [45]. CC from the normal and AD T1-weighted magnetic resonance images have been segmented using Reaction Diffusion (RD) level set method and LB eigen values have been extracted from them. LB eigen values are positive sequence of infinite series that describe the intrinsic geometry of objects. The significant features have been selected based on Information Gain and classification using K-Nearest Neighbour (KNN), Support Vector Machine (SVM) and Naïve Bayes (NB) has been done. KNN is able to give maximum classification accuracy.

There are various ongoing projects in institutes like IIT Madras, IIT Bombay, BMS College of Engineering, etc. based on ADNI [42].

III. Issues and Gaps

The following gaps in current research have been identified based on review of literature:

1. In the real world scenario, patient recordings are discrete and recorded at varying time intervals. Existing models fail to learn continuous time progression model from discrete recordings. The proposed solution will model disease progression as a continuous process. The mined temporal patterns will be continuous in order to make them more interpretable and intuitive.
2. The existing techniques use only patients having a disease to study its progression. The proposed approach will include non case patients as well as case patients in the model.
3. Different patients have different disease progression trajectories. The symptoms and characteristics change at different rates for each individual patient. The proposed method will be able to infer progression of disease in individual patients based on clinical observations.
4. The proposed model will be applied to gain insights into the predictive power of diagnostic and prognostic biomarkers in isolation and in combination for estimation of the probability of developing AD.

IV. Importance of a Machine Learning based Solution in the context of current status

The importance of the proposed approach can be highlighted by listing the benefits of such a solution as follows:

1. Mining of temporal patterns and stages of neurodegenerative diseases vis-à-vis factors and observations will facilitate optimal decision making for clinicians and patients. It will provide doctors with evidence based prediction and recommendation.
2. Early detection of neurodegenerative diseases will lead to reduced burden on patients and healthcare system.

 ISSN: 2454-4248
181 – 187

3. It will aid in improvement of preventive care methods and personalized medicine through prediction of progression of chronic neurodegenerative diseases.
4. Knowledge of temporal patterns related to these diseases can be applied to drug development and clinical trial design.
5. Identification of prognostic biomarkers will help in selection of non demented subjects at risk of developing AD and to test whether treatments have potential to delay onset of AD.
6. Anonymized dataset of acquired clinical, biochemical, imaging data will be open for use by researchers.

V. Proposed Methodology

The methodology broadly corresponds to the CRISP-DM i.e. cross-industry process for data mining framework, which breaks the process of data mining into six major phases namely, business understanding, data understanding, data preparation, modeling, evaluation and deployment. The different phases of proposed methodology are as depicted in the flowchart below.

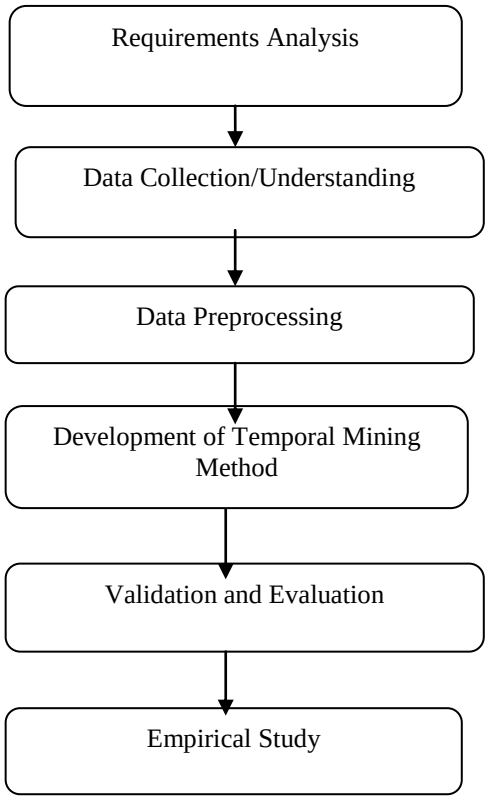

Fig. 1. Flowchart describing phases in the proposed solution

Phase I
Analysis of Requirements
The existing techniques for temporal pattern mining and disease progression modeling will be studied and analyzed. The issues affecting implementation and adoption of these techniques, advantages, disadvantages and limitations of existing techniques will be thoroughly investigated.

Phase II
Data Collection/Understanding
The next step in the design of the approach will be to obtain longitudinal clinical findings of a cohort of patients who are at risk of, or have developed a disease. For each patient, there will be a varying number of set of clinical findings depending on number of visits to the hospital. Each such set of findings will have a temporal attribute associated with it. Some benchmark datasets are available on public repositories like UCI Machine Learning website and Alzheimer's Disease Neuroimaging Initiative. These datasets will be used initially to develop a generic approach for mining temporal information and modeling disease progression.

Phase III
Data Preprocessing
The data collected in previous phase may be noisy, inconsistent and incomplete as is common in real world datasets. Data cleaning techniques like handling missing values, zeroes and data smoothening techniques will be applied to ensure data quality. Data transformation techniques like normalization will be done as tests performed on the same individual by different laboratories can be inconsistent.

Phase IV
Development of machine learning method for temporal mining
An algorithm based on machine learning for mining temporal patterns from irregular data will be developed. The design of the experimental setup will be done. Experiment will be carried out and observations will be recorded. The associated comorbidities will be identified based on pattern matching. The various stages of disease under study will be identified based on these patterns and consultation with domain experts i.e. medical practitioners. Then the longitudinal feature learning will lead to model of the progression of the disease.

Phase V
Validation and Evaluation of Model
The proposed model will be validated by using a k-fold cross validation technique as is typical in data mining literature. Ground truthing will also be done by consulting the medical domain experts in case test dataset is not available. The predictive performance of the proposed approach will be evaluated using measures like AUC, sensitivity and specificity.

Phase VI
Empirical Study
The proposed model will be applied to study the temporal patterns, disease progression and associated factors and comorbidities for Alzheimer's disease.

VI. Conclusion and Future Scope

The challenge of designing a solution to infer temporal patterns for disease progression modeling is an open problem. This paper reviews the literature for disease progression modeling and cognitive disease Alzheimer's Disease. The gaps in current research have been discussed

 ISSN: 2454-4248
181 – 187

and a potential solution based on machine learning is proposed. The outcome of the proposed solution will be to lead to a robust approach for mining temporal patterns from medical data that can be used for building predictive models for Alzheimer's and other neurodegenerative diseases which can lead to cognitive disorders. The approach should be able to scale to large datasets and handle noisy as well as sparse datasets. The approach will have numerous potential applications like assessing the risk scores of a patient developing different disease conditions, understanding disease prognosis, helping in drug development, personalized medicine and optimal decision making. Overall, it will serve as a useful tool to study and model the different stages of progression of neurodegenerative diseases. The methodology for designing such a solution has been proposed.

REFERENCES

[1] C. H. Jackson, L. D. Sharples, S. G. Thompson, S. W. Duffy, and E. Couto, "Multistate markov models for disease progression with classification error", Journal of the Royal Statistical Society: Series D (The Statistician), Vol. 52, No. 2, pp. 193-209, 2003.

[2] K. Ito, S. Ahadieh, B. Corrigan, J. French, T. Fullerton, and T. Tensfeldt, "Disease progression meta-analysis model in alzheimer's disease", Alzheimer's & Dementia, Vol. 6, No. 1, pp. 39-53, 2010.

[3] K. Exarchos, T. Exarchos, C. Bourantas, M. Papafaklis, K. Naka, L. Michalis, O. Parodi, and D. Fotiadis, "Prediction of coronary atherosclerosis progression using dynamic bayesian networks", In IEEE EMBC, pp. 3889-3892, 2013.

[4] W. De Winter, J. DeJongh, T. Post, B. Ploeger, R. Urquhart, I. Moules, D. Eckland, and M. Danhof, "A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying type 2 diabetes mellitus", Journal of pharmacokinetics and pharmacodynamics, Vol. 33 No. 3, pp. 313-343, 2006.

[5] T. M. Post, J. I. Freijer, J. DeJongh, and M. Danhof, "Disease system analysis: basic disease progression models in degenerative disease", Pharmaceutical research, Vol. 22, No. 7, pp. 1038-1049, 2005.

[6] R. Sukkar, E. Katz, Y. Zhang, D. Raunig, and B. T. Wyman, "Disease progression modeling using hidden markov models", In IEEE EMBC, pp. 2845-2848, 2012.

[7] M. Cohen, A. Grossman, D. Morabito, M. M. Knudson, A. Butte, and G. Manley, "Identification of complex metabolic states in critically injured patients using bioinformatic cluster analysis", Critical Care, Vol. 14, No. 1, R10, 2010.

[8] P.D. Yoo, Y. Shwen Ho, N. Jason, M. Charleston5, N. K. Saksena, P. Yang, and A. Y. Zomaya, "Hierarchical kernel mixture models for the prediction of AIDS disease progression using HIV structural gp120 profiles", BMC Genomics, Vol. 11, No. 4, S22, 2010.

[9] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via fused sparse group lasso", In ACM KDD, pp. 1095-1103, 2012.

[10] J. Zhou, J. Liu, V. A. Narayan, and J. Ye, "Modeling disease progression via multi-task learning", NeuroImage, Vol. 78, pp. 233-248, 2013.

[11] J. Yang, J. J. McAuley, J. Leskovec, P. LePendu, and N. Shah, "Finding progression stages in time-evolving event sequences", In WWW, pp. 783-794, 2014.

[12] X. Wang, D. Sontag, and F. Wang, "Unsupervised Learning of Disease Progression Models", In ACM KDD, 2014.

[13] J. Zhou, F. Wang, J. Hu, and J. Ye, "From Micro to Macro: Data Driven Phenotyping by Densification of Longitudinal Electronic Medical Records", In ACM KDD, 2014.

[14] J. Wu, J. Roy, and W. F. Stewart, "Prediction modelling using ehr data: challenges, strategies, and a comparison of machine learning approaches", Med. Care, Vol. 48, No. 6, S106, 2010.

[15] R. Moskovitch, Y. Shahar, "Classification-driven temporal discretization of multivariate time series", Data Mining and Knowledge Discovery, Vol. 29, No. 4, pp. 871-913, 2014. DOI 10.1007/s10618-014-0380-z

[16] J. Hardy, D.J. Selkoe, "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics", Science, Vol. 297, pp. 353–356, 2002.

[17] C.R. Jack Jr, D.S. Knopman, P. S. Aisen, et al. "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade", Lancet Neurol, Vol. 9, pp.119–128, 2010.

[18] C.R. Jack Jr, V.J. , et al. "Serial PIB and MRI in normal, mild cognitive impairment and Alzheimer's disease: implications for sequence of pathological events in Alzheimer's disease", Brain, Vol. 132(Pt 5), pp.1355–1365, 2009.

[19] R.A. Sperling, P. S. Aisen, et al., "Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease", Alzheimers Dement, Vol. 7, pp.280–292, 2011.

[20] Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. Lancet Neurol. 2010; 9:1118–1127.

[21] http://adni.loni.usc.edu/

[22] Trojanowski JQ, Vandeerstichele H, Korecka M, Clark CM, Aisen PS, Petersen RC, et al. Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects. Alzheimers Dement. 2010; 6:230–238.

[23] Shaw LM. PENN biomarker core of the Alzheimer's disease Neuroimaging Initiative. Neurosignals. 2008; 16:19–23.

[24] Aisen PS, Petersen RC, Donohue MC, Gamst A, Raman R, Thomas RG, et al. Clinical core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. Alzheimers Dement. 2010; 6:239–246.

[25] Clark CM, Davatzikos C, Borthakur A, Newberg A, Leight S, Lee VM, et al. Biomarkers for early detection of Alzheimer pathology. Neurosignals. 2008; 16:11–18.

[26] Wu X, Chen K, Yao L, Ayutyanont N, Langbaum JB, Fleisher A, et al. Assessing the reliability to detect cerebral hypometabolism in probable Alzheimer's disease and

amnestic mild cognitive impairment. J Neurosci Methods. 2010; 192:277–285.

[27] Leung KK, Barnes J, Ridgway GR, Bartlett JW, Clarkson MJ, Macdonald K, et al. Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. Neuroimage. 2010; 51:1345–1359.

[28] Hinrichs C, Singh V, Mukherjee L, Xu G, Chung MK, Johnson SC. Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset. Neuroimage. 2009.

[29] Hinrichs C, et al. Predictive markers for AD in a multi-modality framework: an analysis of MCI progression in the ADNI population. Neuroimage. 2011; 55(2):574–89.

[30] Padilla P, et al. NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer's disease. IEEE Trans Med Imaging. 2012; 31(2):207–16.

[31] Zhang D, et al. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011; 55(3):856–67.

[32] Zhang D, Shen D. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. Neuroimage. 2012; 59(2):895–907.

[33] Jack CR Jr, Bernstein MA, Borowski BJ, Gunter JL, Fox NC, Thompson PM. Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative. Alzheimers Dement. 2010; 6:212–220.

[34] Wang H, N F, Huang H, Risacher S, S A, Ding C, Shen L. ADNI*. Sparse Multi-Task Regression and Feature Selection to Identify Brain Imaging Predictors for Memory Performance. IEEE Conference on Computer Vision. 2011:557–562.

[35] Mattila J, et al. A disease state fingerprint for evaluation of Alzheimer's disease. J Alzheimers Dis. 2011; 27(1):163–76.

[36] Soininen H, et al. Software tool for improved prediction of Alzheimer's disease. Neurodegener Dis. 2012; 10(1-4):149–52.

[37] Meda SA, et al. A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the ADNI cohort. Neuroimage. 2012; 60(3):1608–21.

[38] Wang H, et al. Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort. Bioinformatics. 2012; 28(2):229–37.

[39] Ito K, Corrigan B, Zhao Q, French J, Miller R, Soares H, et al. Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. Alzheimers Dement. 2011; 7:151–160.

[40] Samtani MN, et al. An improved model for disease progression in patients from the Alzheimer's disease neuroimaging initiative. J Clin Pharmacol. 2012; 52(5):629–44.

[41] Llano DA, Laforet G, Devanarayan V. Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. Alzheimer Dis Assoc Disord. 2010; 25:73–84.

[42] http://adni.loni.usc.edu/study-design/ongoing-investigations/

[43] K. R. Anandh, C. M. Sujatha, S. Ramakrishnan, "Atrophy analysis of corpus callosum in Alzheimer brain MR images using anisotropic diffusion filtering and level sets", in Proc. of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Aug 2014, pp:1945 – 1948, Chicago.

[44] M. Kayalvizhi, G. Kavitha, C.M. Sujatha, S. Ramakrishnan, "Minkowski functionals based brain to ventricle index for analysis of AD progression in MR images", Measurement, Volume 74, October 2015, pp. 103–112.

[45] Anandh Kilpattu Ramaniharan, Sujatha Chinnaswamy Manoharan, Ramakrishnan Swaminathan, "Laplace Beltrami eigen value based classification of normal and Alzheimer MR images using parametric and non-parametric classifiers", Expert Systems with Applications, Volume 59, 15 October 2016, Pages 208–216.