**Objective:** Prediction of Specialty Diagnostic Procedures for the Patients with Cognitive Impairment Diseases Using Deep Representation Learning of Electronic Health Record

**Significance:** Early detection of cognitive decline may be critical to the efforts to stop dementia progression, including Alzheimer’s disease (AD) and AD-related dementias (ADRD) [1]. Dementia and mild cognitive impairment (MCI) are under-diagnosed even though dementia is one of the major causes of mortality and morbidity in older people worldwide. Early detection of cognitive decline may be critical to the efforts to stop dementia progression [2] A recent research study based on the Medicare data showed that 85% of people first diagnosed with dementia were diagnosed by a non-dementia specialist physician, usually a primary care doctor, and an “unspecified dementia” diagnosis was common [3]. Among those diagnosed by a non-dementia specialist, 33% of patients were given a diagnosis that lacked a specific type of dementia, compared to 22% of patients diagnosed by a specialist. Diagnosis often occurs late in the illness trajectory. Inefficient identification of dementia type leads to low-value care including ineffective and burdensome treatments and procedures.

The NIH has highlighted the need for valid machine learning approaches for mining data from multiple sources and predicting cognitive decline. This proposed pilot study will create a model to help early diagnosis of MCI and AD. The proposed model will identify patients who are likely to have cognitive impairment (Mild Cognitive Impairment and Alzheimer’s Disease) and need timely referral to specialists for further assessment as well as recommending the necessary specialty diagnostic procedures for these patients. The diagnostic procedures may include labs, imaging and cognitive tests. This model not only assists primary care providers in identifying the high risk patients who need further assessment, but can suggest effective diagnostic procedures that help with precise decision making. The proposed automated system enables clinicians to initiate the necessary diagnostic workups with adequate support awaiting in-person specialty visits, based on the data-driven recommended diagnostic procedures.

**Innovation:**

This pilot is responsive to the NIA priorities [2] and includes innovations in representation learning, prediction, and recommendation models. In summary, representation learning helps to represent the high-dimensional complex data such as Electronic Health Record in lower dimensions. We propose an iterative deep representation learning algorithm that simultaneously learns the relevant encounters and the data representation based on the relevant encounters. The prediction and recommendation models are based on an innovative joint KNN graph of the patients and providers with low-dimensional representation of the medical record elements as input. This feasibility study will hopefully yield data that will guide the design of larger clinical studies for earlier diagnosis and assessment of MCI and AD in diverse older adults.

**Approach:**

We propose a recommender system to predict the necessary diagnostic procedures as well as designing an explainable model to measure the effectiveness of these procedures in early and proper order. In our successful preliminary work, we have utilized a method based on ensemble learning to the Endocrine specialty diagnostic orders [4]. Another related work by our team is ClinicNet which uses a feed-forward neural network to automate inpatient electronic order checklists [7].

* **Cohort:** We address the prediction of specialty clinical diagnostic procedures for the outpatients referred to Stanford Health Care Neurology Clinic between Jan 2008 and Dec 2020, diagnosed with cognitive impairment (Mild Cognitive Impairment and Alzheimer’s Disease) identified by the ICD9/ICD10 diagnostic codes.
* **Representation (Embedding) Learning:** Since the feature dimension for each patient is huge, learning low-dimensional feature representations for the patients would be helpful [8]. Similar to the word embedding learning (e.g. Word2Vec [9]), each patient’s low-dimensional representation learning is based on the learned embedding for each of the elements such as medications, lab results, etc. We propose an iterative embedding learning algorithmthat simultaneously learns the relevant encounters and the element embeddings based on the relevant encounters.
* **Recommendation Model:** Embedding-based recommender systems are implemented based on the *k-*Nearest Neighbor (KNN) graph in the literature [10]. In our problem, each diagnostic order depends not only on the patient’s condition but also on the specialist’s preference as well. In the proposed model, the recommendations are based on an innovative joint KNN graph of the patients and providers.
* **Evaluation Metrics:** We will use AUROC, Recall@k, and Precision@k for evaluation. We will compare the proposed method with other baseline methods as well as the clinical guidelines.
* **Missing data and plans to handle it:** Similar to our previous work in [5] the problem of missing data will be addressed using an appropriate encoding of the patients’ previous lab results and procedures. Specifically, each lab result is encoded into a binary vector of size 4, where the first dimension shows whether any previous test results exist for the patient and dimensions 2- 4 respectively correspond to the ‘low’, ‘normal’ and ‘high’ lab results.
* **Anticipated challenges and plans to overcome:** We are aware that no approach would be very accurate [2], but this pilot work is vital in developing and testing a model that would act like a screening test that identifies higher risk patients warranting more diagnostics for intervenable causes. Through this feasibility study, we hope to identify and to overcome this challenge by meticulous labeling for dementia diagnosis in the data and ensuring enough follow-up through claims data and EHR data. Knowledge gained through this pilot will serve as preliminary work for an R21/RO1 application to NIH to facilitate early diagnosis of MCI/AD.

**References**

1. Rasmussen, Jill, and Haya Langerman. "Alzheimer’s disease–why we need early diagnosis." Degenerative neurological and neuromuscular disease 9 (2019): 123.
2. <https://grants.nih.gov/grants/guide/rfa-files/RFA-AG-20-051.html>
3. <https://news.usc.edu/160355/dementia-diagnosis-usc-study-specialty-care/>
4. Morteza Noshad, Ivana Jankovic, Jonathan H Chen, "Clinical Recommender System: Predicting Medical Specialty Diagnostic Choices with Neural Network Ensembles", KDD Data Science for Healthcare workshop (DSHealth 2020)
5. Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
6. Caselles-Dupré, H., Lesaint, F. and Royo-Letelier, J., 2018, September. Word2vec applied to recommendation: Hyperparameters matter. In Proceedings of the 12th ACM Conference on Recommender Systems (pp. 352-356).
7. Wang, J. X., Sullivan, D. K., Wells, A. C. & Chen, J. H. ClinicNet: machine learning for personalized clinical order set recommendations. Jamia Open (2020) doi:10.1093/jamiaopen/ooaa021.
8. Rajkomar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M. and Sundberg, P., 2018. Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), p.18.
9. Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
10. Caselles-Dupré, H., Lesaint, F. and Royo-Letelier, J., 2018, September. Word2vec applied to recommendation: Hyperparameters matter. In Proceedings of the 12th ACM Conference on Recommender Systems (pp. 352-356).

**Key Personnel**:

Our team involves an excellent multi-disciplinary combination of researchers with clinical and computer science backgrounds:

* **Jonathan H. Chen (co-PI)** is an Assistant Professor of Medicine in the Center for Biomedical Informatics Research. He has tremendous expertise in both clinical medicine (he is an internist) and in modeling aspects of clinical decision support systems (he is a bioinformatician) . His unique skills will make this project possible. He has collaborated with Dr. Periyakoil (SAGE PI) in the past and is uniquely poised to complete the current project leveraging the SAGE center infrastructure.
* **Morteza Noshad (co-PI)** is a Postdoctoral Research fellow with computer science background. He will lead the main portion of the work from data processing to model development and evaluation.

**Budget Justification ($30,000):** Postdoc Salary ($20,000), Student Salary ($6,000), Computing Resources ($2,000), Conference/ Publication / Poster Fees ( $2,000)