Interim Progress Report

**1. The research scope**

My research interests lie primarily in the intersection of interpretable deep learning and Alzheimer’s disease assessment. In recent years, deep learning, a form of machine learning, has achieved remarkable success in many applications, including image classification, voice understanding, language translation etc., by directly extracting relevant information from high-dimensional raw data.

Aided by the rapid digitalization of medical records, diagnoses and treatments, the advances of artificial intelligence (AI) for medical applications present many bright opportunities to reshape the future of medicine. More and more AI tools have cleared the regulatory hurdles and have been deployed in real-world clinical settings. However, there still remains a big gap between the uncountable number of research papers in the field of medical AI and the number of systems that have been successfully translated to clinical practice.

The central goal of my research is to achieve a machine learning-based diagnostic tool for dementia assessment using routinely collected data. In the following subsections, I will firstly discuss the burden of Alzheimer’s disease which motivates me the most to conduct my PhD study, and then present a concise introduction of the general applications of machine learning in medicine, and lastly present our efforts towards bridging the gap of machine learning applications between the medical community and the machine learning community.

**1.1 The burden of Alzheimer’s disease**

Worldwide, there are nearly 10 million new cases of dementia annually, of which Alzheimer’s disease (AD) is the most common. As the population ages, the estimated number of Americans who suffer from Alzheimer’s disease is projected to be doubled by 2050. Tremendous efforts from the society have been made to alleviate the Alzheimer’s burden, including the exploration of new clinical measurements for better diagnosis of the disease, drug development for Alzheimer’s treatments, and providing a safe and supportive environment for Alzheimer’s patients.

The increase in clinical demand for timely diagnosis of Alzheimer’s disease will likely contribute to an already considerable burden of morbidity and mortality among the elderly. Due to growing scarcity of clinical dementia specialists, new approaches are needed to improve the diagnosis of cognitive impairment and dementia. Tremendous efforts have been made towards the development of new biomarkers for Alzheimer’s disease, however, these novel diagnostic measures remain limited to research contexts. Thus, developing a reliable diagnostic tool using only the traditional measures as conducted in routine procedures brings many values for aiding clinicians to deliver timely but accurate diagnosis of Alzheimer’s disease.

**1.2 Machine learning for medical applications**

Machine learning algorithms empower computers to solve tasks by learning directly from the data. The advances in medical imaging and the upsurging digitization of medical records have generated an unprecedented amount of medical data which brought many opportunities of transforming hospital care with AI insights.

In the past few years, deep learning has made many breakthroughs in various domains, including imaging recognition, speech understanding, language translation and so on. Unlike the traditional machine learning approaches, deep learning requires very little engineering by hand and usually outperforms the previous generation of machine learning models, especially when used to discover the intrinsic structures of high-dimensional data.

While the popularity of deep learning in medical applications has increased dramatically, the “black-box” nature of many deep learning models caused doubts and concerns, especially when making high-stake operations, of not having transparency in the decision process. Thus, the transition from developing “black-box” deep learning models to designing interpretable models became a trend in the community of medical AI research.

**1.3 Bridging the gaps between the clinical side and machine learning side**

One of the major factors that impact the practicality of such a machine learning system is its diagnostic scope. Many research papers investigated models’ performance under over-simplified scenarios such as a binary classification task between Alzheimer’s disease and healthy control. However, the complexity of the real-world neurologic diagnosis is far beyond this binary setting. We dedicated ourselves to increasing the diagnostic coverage for individuals from different stages of cognitive impairment and for various etiologies of dementia, including Alzheimer’s disease, Parkinson disease, vascular disease etc. When an individual visits a memory clinic, there are no prior assumptions on which diagnostic category that this subject might belong to. Thus, to make a machine learning model clinically relevant, the diagnostic scope that the model can handle plays a critical role. To support the aiming scope, I spent a tremendous amount of time with my colleagues on searching, collecting, and cleaning these unorganized data from multiple cohorts, including publicly available datasets and in-house data from our collaborators.

As mentioned above that interpretability is a necessary component when using machine learning models to make high-stake decisions, one of my research focuses is on developing interpretable deep learning models to maximize our understanding of the model’s decision logic. During my PhD study, I proposed a new method for developing an intrinsically interpretable deep learning model to analyze brain MRI scans. I also systematically applied one of the existing state of the art interpretable methods in a broader scope of dementia assessment. These studies linked the atrophy patterns of the anatomical brain structures with the risk of having declined cognition and dementia, and thus contributed to a better understanding of the disease.

Many machine learning researchers don’t have the priority of working with medical experts while conducting their research. We instead have the access to working closely with medical professionals like neurologists, radiologists, and MD students. We involved them during all stages of the research, including, defining a clinically relevant task for the model to solve, selecting input features based on the readiness of the information in their day-to-day work, evaluating the model against their domain knowledge.

We dived deep into the pathological level of how the disease was defined and then grounded our deep learning framework by rigorously cross validating the model’s interpretable results against neuropathological evidence. In addition to using neuropathological findings to ground the model’s prediction, we also queried diagnosis and ratings directly from neurologists and radiologists to validate the model using their domain knowledge. I believe that these carefully designed clinical-level validation helped relieve, to some extent, the uncertainty and distrust of AI prediction in the medical community.

**2. Research accomplishments**

I published 3 research papers as the first author and am having the fourth paper under review by *Nature Communications*. I briefly described these works below in the chronological order. I also summarized the major contributions of these works to the research community in this field.

**Paper 1:** Qiu, Shangran, et al. "*Fusion of deep learning models of MRI scans, Mini–Mental State Examination, and logical memory test enhances diagnosis of mild cognitive impairment.*" Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10 (2018): 737-749.

**Overview:** This is the first project that I worked on since I joined Prof. Kolachalama’s lab. In this paper, we demonstrated that adding MRI data to other clinical measures, i.e., Mini–Mental State Examination and logical memory test, enhanced the accuracy of detecting the early stage of cognitive impairment. By working on this project, I equipped myself with the basic skills of training and evaluating various types of neural networks, including convolutional neural network and multi-layer perceptron, which laid down the technical fundamentals that I have been using throughout my PhD study.

**Paper 2:** Qiu, Shangran, et al. "*P1‐119: Enhancing deep learning model performance for AD diagnosis using ROI‐based selection.*" Alzheimer's & Dementia 15 (2019): P280-P281.

**Overview:** In this abstract, we investigated the brain regions that are most relevant to the Alzheimer’s disease diagnosis. Specifically, we systematically trained many convolutional neural networks to classify the disease by taking as input different cross-sectional planes from the volumetric brain MRI scans. The region of interests (ROI) was thus determined by observing the distribution of the model’s accuracy over locations from which cross-sectional planes were sampled.

**Paper 3:** Qiu, Shangran, et al. "*Development and validation of an interpretable deep learning framework for Alzheimer’s disease classification.*" Brain 143.6 (2020): 1920-1933.

**Overview:** This is the first major work that we as a team had been collaboratively working on. We presented an interpretable deep learning framework to detect the individuals with Alzheimer’s disease from the healthy control group. In the previous works that we completed, “black-box” neural networks were used to classify disease without knowing any details on how the decision was made. To go beyond simply predicting a disease status, I dedicated myself to designing and developing an intrinsically interpretable deep learning model for Alzheimer’s disease classification.

Inspired by the existing methods of interpretable deep learning by that time, I conceived a novel approach to relate the structural changes in the brain to the degree of risk that the subject might have Alzheimer’s disease. This method is capable of producing high resolution volumetric disease probability maps (DPM) whose voxel value directly represents the probability that the individual might have Alzheimer’s disease. This DPM can be overlaid on the brain MRI scan to reveal the regions with higher risk.

In addition to the method development, we also conducted a set of rigorous validations of the deep learning model against established medical evidence and expert’s domain knowledge. To urge multi-disciplinary collaboration, I worked closely with neurologists and MD students to design and implement domain-specific validations of the model from various angles, including validating model predicted high risk regions using biomarker density as ground truth, and comparing model’s predictive accuracy with a group of neurologists on a subset of subjects.

**Paper 4:** Qiu, Shangran, et al. "*Multimodal deep learning for Alzheimer’s disease dementia assessment.*" Under review by Nature Communications.

**Overview:** In this work, we extended the diagnostic scope of the previously developed interpretable deep learning framework from an over-simplified binary classification task to a more comprehensive multi-task setting to cover most of the individuals who visit memory clinics. First, the framework is designed to distinguish different stages of the cognitive impairment, including the normal cognition stage, mild cognitive impairment stage and the dementia stage. If a case was predicted as demented, the framework can further separate whether the dementia was caused by Alzheimer’s disease or other etiologies, i.e., parkinsonian dementia, vascular dementia etc.

Instead of using the previously developed DPM method, we systematically applied the start of the art SHAP interpretable framework to understand how the model attributed its multi-task predictions to all input signals. SHAP is a post-hoc interpretable framework that can be used to interpret a broad spectrum of machine learning models. Based on the derived interpretable results, we quantified the contributions of each clinical measure to dementia diagnosis and visualized brain atrophy patterns on top of MRI scans by analyzing the SHAP-based saliency map derived from the convolutional neural network.

We again rigorously validated our deep learning model using not only the pathological findings from post-mortem examinations, but also experts’ domain knowledge derived from 17 neurologists and 7 radiologists. The positive correlation between the SHAP-based saliency map and radiologists’ local atrophy rating demonstrated the consistency between the model’s understanding of the brain structures and radiologists’ expertise. By running head-to-head comparisons between the model and experts’ performance on the same subset of data given the same amount of information, we demonstrated that the model operated at the similar accuracy level as averaged experts over several tasks.

**Major contributions:**

1. We set a high standard of validating machine learning models using established medical evidence, for example, the neuropathological measures from post-mortem examinations, and using expert’s domain knowledge, for example, correlating model predicted atrophy rating with radiologist’s rating.
2. We dramatically expanded the diagnostic scope of the deep learning framework and achieved the scope of diagnosing most of the individuals who visit memory clinics.
3. We developed a novel interpretable deep learning method to visualize instance-specific disease risk over MRI, and systematically applied SHAP-based interpretable method in a broader domain of dementia assessment.
4. We rigorously validated the model's performance on many external datasets to comprehensively evaluate the generalizability of the model’s predictive ability when facing never-seen data.

**3. Time scale for completion of the PhD**

I plan to give a department seminar talk before the APS March meeting if all committee members have overlapped availability, and preferably in the upcoming 2 weeks (Feb. 22nd – Mar. 4th).

I plan to give my final defense in May 2022.