**Project Title**: Uncovering heterogeneous cardiometabolic and genomic factors in Alzheimer's Disease etiology through an integrated radiogenomic approach

# Abstract (300 Words max)

Alzheimer's disease (AD) is a multifactorial, heterogeneous neurodegenerative disease1. Cardiometabolic risk factors play critical roles in AD pathogenesis and clinical manifestation through pathways that may be dependent or independent of genetic factors. Specifically, multi-faceted cardiometabolic factors might contribute to heterogenous phenotypes of AD neurodegeneration and proteinopathy pattern in the brain. However, the synergic effects of multiple cardiometabolic factors that contribute to distinctive AD subtypes have not been fully investigated. Moreover, genomics and neuroimaging data provide complementary information on disease mechanisms. An integrated radiogenomic approach would enable a comprehensive understanding of the genotype-phenotype interaction to potentially inform more personalized clinical intervention. However, both neuroimaging and genomic data are large in dimension, and genomic effects on AD pathology are complex and might occur at multiple genomic loci concurrently with complex gene-gene interactions, involving both risk and protective effects. There is a lack of effective and efficient biologically-driven analytical approaches to extract high-level AD-related multi-genome interactions from the whole-genome sequence to infer genotype-phenotype interaction.

In this proposed project, we **hypothesize** that: **a)** synergic composited effect of cardiometabolic conditions contributes to the heterogeneity of AD-related disease pathophysiological patterns in the brain; **b)** the risk of AD is characterized by joint genotype-phenotype interactions, with phenotype manifested from both brain pathophysiological patterns and metabolic factors. We propose to test these hypotheses through a data-driven integrated radiogenomic approach with two **Specific Aims:** **1)** to evaluate the multi-faceted cardiometabolic contributions to the heterogeneity of AD pathophysiological progression; **2)** to derive the joint radiogenomic and cardiometabolic patterns associated with Alzheimer's Dementia etiology.

Completing these objectives in this project will lay the foundational work for future extramural grant project directions. The results on genotype-phenotype interaction and multi-faceted metabolic factors contribute to the heterogeneous AD etiology will generate new insights for potential therapeutic targets. **b)** The radiogenomic approach and multi-modal joint feature embedding framework are translatable to more general multi-morbidity conditions.

## Submitting Investigator, Co-Investigators, and other Key Personnel information

PI Dr. **Da Ma** is an Assistant Professor of Gerontology and Geriatric Medicine with research expertise in neuroimage genomics and machine learning for Alzheimer's Disease. Dr. **Suzanne Craft** is the Director of the WF Alzheimer'ss Disease Research Center (ADRC) and is a national leader in AD mechanisms and therapeutic approaches. Dr. **Tom Register** is the leader of the Fluid Biomarker Service in the Neuropathology Core of the WF ADRC and an expert in multidisciplinary, translational studies in human and nonhuman primates. **Metin Gurcan** is the Director of the Wake Forest Center for Biomedical Informatics and is an expert in medical and clinical image-based informatics analysis. Dr. **Sam Lockhart** is an Assistant Professor of Gerontology and Geriatric Medicine and is the co-leader of the Imaging Biomarker Core at WF ADRC. Dr. **James Bateman** is an Assistant Professor of Neurology, behavioral neurologist, and the co-leader of the Clinical Core at WF ADRC. Dr. **Byron C. Jaeger** is an Assistant Professor of Biostatistics and Data Science in the Division of Public Health Sciences, with research expertise in machine learning, statistical computing, and computational optimization.

## Suggested Reviewers – cannot be a current or previous mentor or co-author on a publication within the last 2 years

## Specific Aims

Alzheimer's disease (AD) is a multifactorial and heterogeneous neurodegenerative disease1. Both genetic and cardiometabolic factors play crucial roles in disease etiology and progression2,3. AD patients with selected metabolic dysfunctions and genomic risk factors may experience different rates of cognitive decline, distinctive patterns of neurodegeneration, and varying ages of dementia onset. Specifically, cardiometabolic factors, such as hyperglycemia, hypertension, high blood pressure, and high blood glucose leve, affect the potential risk of dementia incidence through complex and composite causal pathways either dependent2 or independent3 of genetic risk factors. The proposed project aims to use data-driven approaches to (i) investigate the multi-faceted cardiometabolic contributions towards the heterogeneity of AD pathophysiological patterns in the brain; and (ii) construct joint radiogenomic and cardiometabolic patterns associated with Alzheimer's Dementia etiology.

Metabolism is crucial for brain function and health4. Metabolic disorders play vital roles in AD onset and progression5, affecting brain insulin resistance4,6, cognitive function7, cerebral vascular abnormalities4,8, neurodegeneration9, as well as amyloid accumulation10. Moreover, cardiometabolic factors are modifiable risk factors, possessing great potential for clinical interventions to reduce dementia risk. However, **both cardiometabolic diseases and AD are multi-faceted and have heterogeneous phenotypes**11, with each phenotype corresponding to distinctive clinical conditions and/or genetic risk factors12–14,15,16. Uncovering the synergic effect of multiple cardiometabolic factors that contribute to distinctive AD progression patterns in the brain possesses a great translational potential to facilitate the development of effective interventions that target specific factors that maximize personalized treatment benefit.

Genomic risk factors play an important causal role in both AD and cardiometabolic disorder etiology. Recent genome-wide association studies (GWAS) identified multiple shared common genetic loci that are associated with both AD and cardiometabolic traits17,18, indicating potential synergic genetic relationships of the two health conditions. However, genomic effects on disease etiology are complex, involving multiple genomic loci with complex gene-gene interactions. **Effective analysis at the whole genome level may provide greater disease risk predictive power compared to traditional single-genetic-loci-based analysis**. Furthermore, genomics and neuroimaging data provide complementary information on disease etiology. Specifically, our recent study19 revealed differential contributions to AD risk at different disease stages. Therefore, it is of great interest todevelop **effective multi-modal radiogenomic approaches to achieve accurate prediction of disease risk,** incorporating both genomic and cardiometabolic factors.

The **overarching objective** of this study is to uncover the shared and interactive genotype-to-phenotype pathways that induced the common etiology and multi-faceted sub-phenotypes for Alzheimer's Disease and cardiometabolic dysfunction through an integrated radiogenomic approach. We **hypothesize** that: **a)** Various cardiometabolic conditions contribute to the heterogeneity of AD-related disease pathophysiological patterns in the brain; **b)** The risk of AD is characterized by joint genotype-phenotype interactions, with phenotype manifested from both brain pathophysiological patterns and metabolic factors. We propose to test these hypotheses through a data-driven integrated radiogenomic approach with two **Specific Aims**.

**Aim 1: Study the multi-faceted cardiometabolic contribution to the heterogeneity or AD pathophysiological pattern**. We will first identify different sub-phenotypes of AD **pathophysiological** patterns using multi-modal neuroimaging data using semi-supervised clustering methods. We will then construct multivariate machine-learning-based classification models with posthoc feature importance analysis to derive the interactive cardiometabolic factors that contribute to different AD subtypes.

**Aim 2: Derive the joint radiogenomic and cardiometabolic patterns associated with Alzheimer's Dementia etiology**. We will first derive a multi-modal joint feature embedding framework through self-supervised representation learning to achieve generalizable dimension reduction for genomic and neuroimage data. We will then construct multi-modal supervised deep-learning-based predictive models for future risk of AD onset, and use explainable AI methods to derive joint radiogenomic and cardiometabolic patterns.

Impact: The proposed development project will lay the groundwork for numerous future directions. **a)** The results on genotype-phenotype interaction and multi-faceted metabolic factors towards heterogeneous AD etiology will generate new insights for potential therapeutic targets. **b)** The radiogenomic approach and multi-modal joint feature embedding framework are translatable to more general multi-morbidity conditions. We believe this project will generate preliminary data and novel findings for extramural R01 grant applications.

## Background and Significance

*Translational:* AD is the most prevalent dementia type with complex and multifactorial etiology, and effective intervention strategies is limited. Cardiometabolic factors are one of the few modifiable risk factors that could reduce the risk of dementia. The proposed research will contribute to the missing fundamental understanding of the synergic contribution among various cardiometabolic factors towards different sub-phenotypes of AD. Furthermore, the identification of specific genomic signatures that are common for the development of cardiometabolic dysfunction and AD presents opportunities to discover novel biological mechanisms and provide directions for personalized therapeutic opportunities. The proposed radiogenomic approach will generate new knowledge about the integral joint effect of genomic, neuroimage, and cardiometabolic factors on AD etiology. The implication of the proposed study also holds great potential to shed light on the connections and interactions among multiple aging-related chronic diseases at both the molecular level and the pathological level. This has strong clinical implications for patient-specific diagnosis, risk prevention, and intervention planning to improve clinical outcomes for the elderly population, which have a higher risk of developing multi-morbidity. The successful implementation of the project will lead to the potential extramural funded larger project for prospective studies at Wake Forest University School of Medicine to recruit a focused population with both prediabetic and mild cognitive impairment comorbidity conditions to validate the genotype-phenotype interactions, as well as evaluate the effect of potential interventions targeting such comorbidity conditions. *Methodological:* the proposed effective radiogenomic feature embedding framework is generalizable to achieve simultaneous extraction of high-dimensional neuroimage and multi-omic data. The proposed multi-task learning framework is also generalizable to chronic diseases with comorbidity conditions. The semi-supervised clustering framework for disease sub-phenotyping developed in this study would also be generalizable to other age-related chronic diseases with multi-morbidity conditions.

## Innovation

The project brings the following innovations:

1. The neuroimage-derived clustering algorithm for AD sub-phenotyping, in combination with the explainable machine learning approach, enables the detailed exploration of the joint contribution of multiple cardiometabolic factors to AD pathogenesis.
2. The deep-learning-based dimension reduction and feature extraction model investigates the genomic factor involving high-level non-linear gene-gene interaction, in contrast to the traditional GWAS approach in which genomic risk factors were in a single-gene level.
3. The radiogenomic approach using multi-modal deep-learning with joint feature embedding reveals the joint genotype-phenotype effect on the disease pathology.

## Study Team, Investigators, and key personnel

PI Dr. **Da Ma** is an Assistant Professor of Gerontology and Geriatric Medicine with research expertise in neuroimage genomics and machine learning for Alzheimer's Disease. Dr. **Suzanne Craft** is the Director of the WF Alzheimer's Disease Research Center (ADRC) and is a national leader in AD mechanisms and therapeutic approaches. Dr. **Tom Register** is the leader of the Fluid Biomarker Service in the Neuropathology Core of the WF ADRC and an expert in multidisciplinary, translational studies in human and nonhuman primates. **Metin Gurcan** is the Director of the Wake Forest Center for Biomedical Informatics and is an expert in medical and clinical image-based informatics analysis. Dr. **Sam Lockhart** is an Assistant Professor of Gerontology and Geriatric Medicine and is the co-leader of the Imaging Biomarker Core at WF ADRC. Dr. **James Bateman** is an Assistant Professor of Neurology, behavioral neurologist, and the co-leader of the Clinical Core at WF ADRC. Dr. **Byron C. Jaeger** is an Assistant Professor of Biostatistics and Data Science in the Division of Public Health Sciences, with research expertise in machine learning, statistical computing, and computational optimization.

## Approach, Experimental Design, and Methods

In this pilot project, our multidisciplinary team will develop and validate data-driven machine-learning models to uncover the shared and interactive genotype-phenotype pathways that induced the common etiology for Alzheimer's Disease and cardiometabolic abnormality through an integrated radiogenomic approach. We will incorporate all subjects in the community-dwelling cohort from Wake Forest Alzheimer's Disease Research Center (**WFADRC**). In addition, for the neuroimage-based AD subtyping in step 1 of Aim 1 and semi-supervised radiogenomic feature extraction in step 1 of Aim 2, we will also use the **ADNI** dataset to derive neuroimage-genomic features from a larger independent dataset.

**Wake Forest ADRC** cohort**:** The Wake Forest ADRC has collected multi-modal neuroimaging data (MRI + PET), cognitive tests, consensus diagnosis, and other biomarkers, such as CSF and plasma biomarkers, from 657 subjects since 2016. The currently available data includes a total of 284 NC subjects (out of which 25 converted to MCI), 135 MCI subjects (out of which 21 converted to AD), and 45 AD subjects. The proposed project will also aim to support the ongoing continuous efforts to collect genomic data (SNPs) for all the participants. **ADNI** cohort**:** The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a multi-year public-private partnership to test whether serial MRI, PET, clinical and neuropsychological assessment can be combined to measure the progression of amnestic mild cognitive impairment (MCI), which also includes data from NACC. A total of 1893 subjects (626 normal controls, 869 MCI, 398 AD), ranging in age from 55 to 90 years, were recruited from 59 sites across the U.S. and Canada. The participants in the ADNI study were also genotyped (62,901 SNPs in ADNI-1 and 730,525 SNPs in ADNI-2). All genotyping information will be downloaded from the LONI Image Data Archive.

## Analysis plan

### Aim 1: Study the multi-faceted cardiometabolic contribution to the heterogeneity of AD pathophysiology

#### Introduction, background, and rationale:

A significant challenge in developing effective intervention strategies to treat AD is our limited understanding of the relationship and interaction of multiple risk factors in a systematic way. Specifically, metabolic factors are modifiable risk factors that play vital roles in AD etiology and can be potential therapeutic targets for disease-modifying treatment. Metabolism is crucial for maintaining brain health4, and cardiometabolic syndromes such as Type 2 diabetes mellitus (T2DM) might have shared causal pathways with AD 20,21. Cardiometabolic dysfunctions in the body, including hyperglycemia, hypertension, and high blood pressure, are associated with an increased risk of early brain alteration9. On the other hand, the phenotypes and clinical manifestation of cardiometabolic dysfunctions and AD are both heterogeneous11,12–14,15,16, making it challenging to understand the clear pathways for disease etiology. Thus, the *objective of this aim* is to derive the pattern of multi-faceted cardiometabolic contribution towards the heterogeneity or AD pathophysiological progression. To achieve our objective, we will test the working *hypothesis* that the synergic composited effect of various cardiometabolic conditions contributes to the heterogeneity of AD-related disease progression patterns. Our *approach* includes first identifying different sub-phenotypes of AD progression patterns using multi-modal neuroimage using semi-supervised clustering methods, followed by the construction of multivariate machine-learning-based classification models with posthoc feature importance analysis to derive the interactive cardiometabolic factors that contribute to different AD subtypes. The *rationale* for this aim is that the successful implementation of the proposed research will contribute to the current knowledge gap about the synergic contribution of multiple cardiometabolic factors towards different sub-phenotypes of AD-related pathological progression patterns in the brain. The developed semi-supervised clustering framework for disease sub-phenotyping and explainable AI approach would also be generalizable to other age-related chronic diseases with multi-morbidity conditions. Our *expected outcome* is that, upon the completion of the proposed study for Aim 1, we will generate explainable AI models that reveal the relative importance of various cardiometabolic factors contributing to the AD pathophysiological sub-phenotypes, which in term was constructed through semi-supervised deep learning models with multi-modal neuroimage.

#### Justification, Feasibility, and Preliminary Data

Recent studies have revealed heterogenous sub-phenotypes for both metabolic dysfunctions12–14 and AD15,16. The potential synergic pathogenesis mechanism11 leads to the important clinical implication but unfilled research question for this aim. In our previous work, we have developed neuroimaging-based machine-learning models 19,22–28 to differentiate dementia subtypes, as well as identify subgroups AD patients with distinctive longitudinal neurodegeneration progression patterns, which will be furture developed to identify AD subtypes through semi-supervised deep learning approach, and used as the outcome measurement of the research in this aim. The participants in the WFADRC cohort consist of a community-dwelling population at-risk of both cardiometabolic abnormality and Alzheimer's Disease, and underwent comprehensive cardiometabolic profiling, multi-modal neuroimaging, as well as multi-domain cognitive test. This provides the ideal opportunity for investigating the proposed research with the aim of unveiling the multi-faceted conditional relationship of the pathogenesis between these two chronic health conditions through data-driven approaches.

#### Research Design, Methods, and Analysis:

In this aim, we will use a step-wise machine-learning approach to derive cardiometabolic signatures for distinctive AD pathophysiological subtypes in the brain (**Figure 1**). In *step 1* of this aim, we will disentangle the neuropathological heterogeneity to derive distinctive AD subtypes through semi-supervised clustering algorithms expanded from our previously developed generalizable dementia subtyping algorithm22 using generative adversarial network (GAN) with cluster-based loss. The AD subtyping model will be trained and derived from the subjects in the ADNI dataset who were diagnosed as MCI or AD at their baseline visit, and then applied to the WFADRC neuroimage dataset. We will derive different sub-phenotypes based on distinctive neuroimage modalities. 1) the subjects will be clustered into three subtypes based on brain atrophy patterns using T1 structural MRI - medial temporal, parietal, and diffusion subtypes29,30. 2) Using β-Amyloid PET data – frontal, parietal, and occipital subtypes31 3) With the available tau-PET brain image data, we will also derive four distinctive subtypes of tau-deposition patterns: limbic-predominant, medial temporal lobe-sparing patterns, and posterior and lateral temporal patterns resembling atypical clinical variants of AD32. The brain regions used to define the AD-related physio-pathological patterns will be derived from the whole-brain segmentation on T1 structural MRI using the FreeSurfer. The PET images from the same subjects will be registered to the T1 MRI space to derive the corresponding regional PET standardized update value ratio (SUVR). Finally, we will also classify subjects based on the biological definition of the AD-related A/T/N (Amyloid/Tau/Neurodegeneration) subtypes33. The A/T status will be derived from the CSF biomarker, given its relative availability compared to the PET image. The N (neurodegeneration) score will be determined from our previously developed machine-learning-based harmonized AD prediction model 23,34. The harmonized cutoff threshold will be determined from the data by finding the maximum Youden index for each biomarker (i.e. A/T/N) to predict the AD onset.

In *step 2*, we will evaluate the cardiometabolic profiles for each of the AD neurodegenerative subtypes and identify the distinctive signature of composited cardiometabolic factors that correspond to each of the AD subtypes. The cardiometabolic factors measured in the WFADRC include: two-hour dynamic fasting blood glucose (OGTT) and insulin level, systolic blood pressure (SysBP), heart rate (HR), triglyceride, hip-to-waist ratio, body-to-mass index (BMI), cholesterol level (both low-/high-density lipoprotein HDL/LDL), and hemoglobin A1C. To understand the synergic and potentially non-linear concurrent effect of different metabolic factors, we will use machine-learning-based multivariate classification models in combination with the posthoc feature-importance test. Two models will be used: Random-Forest, and Multi-layer perceptron (MLP) 35. Feature importance analysis will be performed to evaluate the relative contribution of each cardiometabolic risk factors to the AD subtypes. We will train a machine learning model with multi-task learning 36 to classify multi-modal AD sub-phenotyping within a single composited framework. This multi-task learning strategy will reveal the synergic cardiometabolic-factor-derived multi-faceted brain pathophysiological patterns.

#### Anticipated Results

The anticipated outcome of step 1 is semi-supervised deep learning models that cluster distinctive AD subtypes based on distinctive patterns with multi-modal neuroimage. Step 2 will generate explainable AI models that reveal the relative importance of various cardiometabolic factors contributing to the AD pathophysiological sub-phenotypes.

#### Potential Pitfalls and Alternative Strategies

In step 1, we anticipate the GAN-based clustering algorithm will provide the most generalizable AD subtypes. In cases our GAN-based approach derives suboptimal AD subtypes, we will use the currently widely-adopted disease subtyping algorithm SuStain (Subtype and Stage Inference) as alternative model37,38. If multi-task learning in step 2 yield suboptimal results, We will perform the brain-wise association study (BWAS) to identify the effect of individual cardiometabolic factor for each of the brain region, followed with the analysis using normalized cardiometabolic index (CMI), calculated as the mean of the z-score across all the cardiometabolic factors, which is a modified version of the original CM 39.

### Aim 2: derive the joint radiogenomic and cardiometabolic patterns associated with AD etiology

#### Introduction, background, and rationale:

Genomic risk factors are essential for both AD and cardiometabolic disorders. However, both genomic and neuroimage data are high-dimensional biomedical data with distinctive data characteristics, making it challenging to analyze at the whole-genome and whole-brain levels. Previous GWAS studies have revealed shared common genomic factors through single loci analysis17,18, indicating potential shared disease etiology. Deep-learning-based approaches showed the potential to learn the joint patterns from both neuroimage and genomic data using multi-modal representation learning 40,41. Our recent study showed that genomic and neuroimage data provide complementary genotype-phenotype information on disease etiology at different stages19,42. The *objective of this aim* is to derive the joint radiogenomic and cardiometabolic patterns associated with Alzheimer's Dementia etiology. To achieve this objective, we will test the *working hypothesis* that the risk of AD is characterized by mutual genotype-phenotype interactions, in which phenotypes are manifested from both brain pathophysiological patterns and metabolic factors. Our *approaches* first extract radiogenomic features through semi-supervised dimension reduction and joint feature embedding. We will then train a multi-modal deep-learning framework to predict the risk of developing AD through non-linear interaction of radiogenic-metabolic factors. The AD-risk-related multi-modal shared patterns will be derived through model explainability post hoc analysis. The *rationale* for this aim is that the proposed research contributes to our understanding of how the integration of genomic factors, neuroimage factors, and cardiometabolic factors work together to contribute to the risk of AD. Our *expected outcome* is that, with the completion of the proposed study, we will construct an effective radiogenomic feature embedding framework to simultaneously extract high-dimensional neuroimage and genomic data, along with an explainable multi-modal prediction model for future risk and survival rate of AD considering multiple genomic and phenotypical risk factors along the disease etiology.

#### Justification, Feasibility, and Preliminary Data

Chronic diseases such as AD is affected by the interaction of genomic, environmental, and lifestyle risk factors. The integrated genotype-phenotype analysis would allow accurate disease risk prediction as well as identification of joint effects across multiple types of risk factors. In our recent works19,42, using multi-modal feature selection on neuroimage and genomic data along with deep-learning-based survival analysis, we unveiled distinctive radiogenomic associations in patient groups at different stages of the disease spectrum. Our results demonstrated that while neuroimaging-based neuropathological biomarkers are better at predicting the onset of dementia of Alzheimer's type for patients with mild cognitive impairment (MCI), genomic information is superior in predicting the risk of dementia for the clinically normal population without identifiable neuroimage biomarkers. The proposed research with this aim will build upon these previous developments and findings, and further extend the radiogenomic approach with semi-supervised multi-type feature embedding and fusion, as well as integrate cardiometabolic risk factors in the disease prediction model.

#### Research Design, Methods, and Analysis:

To achieve the perspective of this aim, we will develop a semi-supervised radiogenomic framework to extract the neuroimage-genomics features that are associated with the joint risk of cardiometabolic dysfunctions and Alzheimer's Disease (**Figure 2**). We will extend our previous work for Neuroimage-genomic Alzheimer's Disease prediction19,42. Two major improvements will be developed over our existing framework: 1) the radiogenomic multi-type features feature extraction and dimension reduction will be achieved through a semi-supervised representation learning approach, instead of our original feature selection methods, to achieve more representative and generalizable latent features in the embedded space. 2) we will include cardiometabolic factors as additional feature vectors in the multi-type feature for joint prediction of AD risk and survival rate, which will enable the identification of distinctive cardiometabolic risk factors that interact with radiogenomic factors in the AD etiology and disease progression.

In *step 1*, we will use a semi-supervised approach to achieve dimension reduction and feature extraction from both data types. Early studies attempted to group the genetic segments – the polymorphisms (SNP) – into ranked functional pathways using prior knowledge of gene-gene interactions43. In this step, deep-neural-network-based variational auto-encoder models will be used to train to represent data with a reduced set of latent variables through self-supervised learning. Autoencoder is an information compression method to project high-dimensional data into representative low-dimension representation, in which an encoder-decoder DNN architecture is trained to recover the original full neuroimage/genomic feature using compressed feature represenatation44,45–49. Specifically, we will use two-dimensional (2D) convolutional auto-encoder architecture for neuroimage data and one-dimensional (1D) convolutional auto-encoder architecture for its advantage in capturing the sequential nature of the SNP-based genomic information45. Since this is a semi-supervised step, no outcome labels were required. We will use the ADNI dataset, which is a longitudinal cohort with a larger sample size that also contains both neuroimage and genomic data for each participant.

In *step 2*, we will use the multi-modal deep-learning framework to joint-learn the risk of developing Alzheimer's Disease through non-linear interactions of radiogenomic factors. The embedded neuroimage and genomic features that are derived from the semi-supervised autoencoder network in the first step will be concatenated factors as joint radiogenomic features, together with cardiometabolic factors as well as demographic and CSF, to feed into a multi-layer-perceptron, which is a fully connected deep neural network. The pre-trained model weights in the joint feature embedding part will be frozen, and the outcome of this second-level network will provide the prediction for dementia onset. The Alzheimer's Disease status will be derived from the adjudicated consensus diagnosis outcome. We will then use the explainable AI methods to derive the integrated radiogenomic patterns that are associated with cardiometabolic dysfunction and AD. Recent developments of explainable AI methods such as occlusion and permutation mapping integrated gradient50, and gradient-weighted class activation map (Grad-CAM) 51 has enabled the feasibility to derive insights about localized and joint importance neurological and genomic feature to predict task of interest. These explainable AI methods will be used to derive important multi-modal features in both the neuroimage genomic data showing associated localized brain regions and genomic loci that are jointly associated with metabolic dysfunction and Alzheimer's Disease.

#### Anticipated Results

We anticipate the successful completion of step 1 of this aim will generate semi-supervised parallel radiogenomic feature embedding models to capture both the genomic and neuroimage patterns from the original high-dimensional space. The completion of step 2 will result in an improved AD risk and survival rate prediction model that effectively takes into account multi-domain genetic and phenotype risk factors.

#### Potential Pitfalls and Alternative Strategies

We expect the proposed study will be feasible. However, if the concatenation of cardiometabolic factors fails to improve AD risk prediction, we will construct an alternative multi-task learning model to use the joint-embedded neuroimage-genomic features to simultaneously predict cardiometabolic dysfunction and AD onset. The resulting radiogenomic pattern will then reflect the common etiology for both comorbidities. The cardiometabolic dysfunction will be derived as the cardiometabolic index (CMI), which is an aggregated measurement of cardiometabolic function 39. If the proposed joint feature embedding framework produces suboptimal multi-modal feature representations due to the limitation of sample size, we will use the multi-modal feature selection approach that we have recently developed to achieve dimension reduction for high-dimensional neuroimaging genomic features24. Mean neuroimage measurements (structural volume, cortical thickness, mean Aβ/tau SUVR update) for each FreeSurfer-segmented brain region will be used as anatomically driven input feature

## Project Timeline - Quarterly milestones and anticipated outcomes with timeline

**Aim 1 Anticipated Outcomes**: multi-faceted cardiometabolic patterns towards AD subtypes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Aim | Milestone | M1-3 | M4-6 | M7-9 | M10-12 |
| 1 | Neuroimage-based AD subtyping | X | X |  |  |
|  | Cardiometabolic profile for AD subtyping |  | X |  |  |
|  | Journal conference publication |  |  | X |  |

**Aim 2 Anticipated Outcomes**: joint radiogenomic and cardiometabolic risk pattern for AD etiology

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Aim | Milestone | M1-3 | M4-6 | M7-9 | M10-12 |
| 2 | Genomic data collection | X | X |  |  |
|  | Genomic feature embedding |  | X |  |  |
|  | Neuroimage feature embedding |  |  | X |  |
|  | Multi-modal explainable AD risk prediction |  |  | X | X |
|  | Journal conference publication, grant submission |  |  | X | X |

# References

1. Lam B, Masellis M, Freedman M, Stuss DT, Black SE. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. Alzheimers Res Ther. 2013 Jan 9;5(1):1. PMCID: PMC3580331

2. Pasqualetti G, Thayanandan T, Edison P. Influence of genetic and cardiometabolic risk factors in Alzheimer's disease. Ageing Res Rev. 2022 Nov 1;81:101723.

3. Tai XY, Veldsman M, Lyall DM, Littlejohns TJ, Langa KM, Husain M, Ranson J, Llewellyn DJ. Cardiometabolic , genetic risk, and dementia: a prospective cohort study. Lancet Healthy Longev. Elsevier; 2022 Jun 1;3(6):e428–e436. PMCID: PMC9184258

4. Craft S. The Role of Metabolic Disorders in Alzheimer Disease and Vascular Dementia: Two Roads Converged. Arch Neurol. 2009 Mar 1;66(3):300–305. PMCID: PMC2717716

5. Kang S, Lee Y ho, Lee JE. Metabolism-Centric Overview of the Pathogenesis of Alzheimer's Disease. Yonsei Med J. 2017 May 1;58(3):479–488. PMCID: PMC5368131

6. Monte SM de la. Brain Insulin Resistance and Deficiency as Therapeutic Targets in Alzheimers Disease. Curr Alzheimer Res. 2012;9(1):35–66.

7. Dove A, Marseglia A, Shang Y, Grande G, Vetrano DL, Laukka EJ, Fratiglioni L, Xu W. Cardiometabolic accelerates cognitive decline and dementia progression. Alzheimers Dement [Internet]. [cited 2022 Nov 30];n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12708

8. Marseglia A, Fratiglioni L, Kalpouzos G, Wang R, Bäckman L, Xu W. Prediabetes and diabetes accelerate cognitive decline and predict microvascular lesions: A population-based cohort study. Alzheimers Dement. John Wiley & Sons, Ltd; 2019 Jan;15(1):25–33.

9. Vergoossen LWM, Jansen JFA, Backes WH, Schram MT. Cardiometabolic determinants of early and advanced brain alterations: Insights from conventional and novel MRI techniques. Neurosci Biobehav Rev. 2020 Aug 1;115:308–320.

10. Tondo G, Boccalini C, Vanoli EG, Presotto L, Muscio C, Ciullo V, Banaj N, Piras F, Filippini G, Tiraboschi P, Tagliavini F, Frisoni GB, Cappa SF, Spalletta G, Perani D, Project on behalf of the NA. Brain Metabolism and Amyloid Load in Individuals With Subjective Cognitive Decline or Pre–Mild Cognitive Impairment. Neurology. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2022 Jul 19;99(3):e258–e269. PMCID: PMC9302934

11. Celis-Morales CA, Franzén S, Eeg-Olofsson K, Nauclér E, Svensson AM, Gudbjornsdottir S, Eliasson B, Sattar N. Type 2 Diabetes, Glycemic Control, and Their Association With Dementia and Its Major Subtypes: Findings From the Swedish National Diabetes Register. Diabetes Care. 2022 Jan 25;45(3):634–641.

12. Wesolowska-Andersen A, Brorsson CA, Bizzotto R, Mari A, Tura A, Koivula R, Mahajan A, Vinuela A, Tajes JF, Sharma S, Haid M, Prehn C, Artati A, Hong MG, Musholt PB, Kurbasic A, De Masi F, Tsirigos K, Pedersen HK, Gudmundsdottir V, Thomas CE, Banasik K, Jennison C, Jones A, Kennedy G, Bell J, Thomas L, Frost G, Thomsen H, Allin K, Hansen TH, Vestergaard H, Hansen T, Rutters F, Elders P, t'Hart L, Bonnefond A, Canouil M, Brage S, Kokkola T, Heggie A, McEvoy D, Hattersley A, McDonald T, Teare H, Ridderstrale M, Walker M, Forgie I, Giordano GN, Froguel P, Pavo I, Ruetten H, Pedersen O, Dermitzakis E, Franks PW, Schwenk JM, Adamski J, Pearson E, McCarthy MI, Brunak S. Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity in newly diagnosed individuals: An IMI DIRECT study. Cell Rep Med. 2022 Jan 18;3(1):100477. PMCID: PMC8784706

13. Nair ATN, Wesolowska-Andersen A, Brorsson C, Rajendrakumar AL, Hapca S, Gan S, Dawed AY, Donnelly LA, McCrimmon R, Doney ASF, Palmer CNA, Mohan V, Anjana RM, Hattersley AT, Dennis JM, Pearson ER. Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. Nat Med. 2022 May;28(5):982–988.

14. Wagner R, Heni M, Tabák AG, Machann J, Schick F, Randrianarisoa E, Hrabě de Angelis M, Birkenfeld AL, Stefan N, Peter A, Häring HU, Fritsche A. Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med. 2021 Jan;27(1):49–57.

15. Risacher SL, Anderson WH, Charil A, Castelluccio PF, Shcherbinin S, Saykin AJ, Schwarz AJ, Initiative F the ADN. Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline. Neurology. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2017 Nov 21;89(21):2176–2186. PMCID: PMC5696639

16. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol. 2011 Sep 1;10(9):785–796. PMCID: PMC3175379

17. Broce IJ, Tan CH, Fan CC, Jansen I, Savage JE, Witoelar A, Wen N, Hess CP, Dillon WP, Glastonbury CM, Glymour M, Yokoyama JS, Elahi FM, Rabinovici GD, Miller BL, Mormino EC, Sperling RA, Bennett DA, McEvoy LK, Brewer JB, Feldman HH, Hyman BT, Pericak-Vance M, Haines JL, Farrer LA, Mayeux R, Schellenberg GD, Yaffe K, Sugrue LP, Dale AM, Posthuma D, Andreassen OA, Karch CM, Desikan RS. Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. Acta Neuropathol (Berl). 2019 Feb 1;137(2):209–226. PMCID: PMC6358498

18. Bone WP, Siewert KM, Jha A, Klarin D, Damrauer SM, Ballas ZK, Bhushan S, Boyko EJ, Cohen DM, Concato J, Constans JI, Dellitalia LJ, Fayad JM, Fernando RS, Florez HJ, Gaddy MA, Gappy SS, Gibson G, Godschalk M, Greco JA, Gupta S, Gutierrez S, Hammer KD, Hamner MB, Harley JB, Hung AM, Huq M, Hurley RA, Iruvanti PR, Ivins DJ, Jacono FJ, Jhala DN, Kaminsky LS, Kinlay S, Klein JB, Liangpunsakul S, Lichy JH, Mastorides SM, Mathew RO, Mattocks KM, McArdle R, Meyer PN, Meyer LJ, Moorman JP, Morgan TR, Murdoch M, Nguyen XMT, Okusaga OO, Oursler KAK, Ratcliffe NR, Rauchman MI, Robey RB, Ross GW, Servatius RJ, Sharma SC, Sherman SE, Sonel E, Sriram P, Stapley T, Striker RT, Tandon N, Villareal G, Wallbom AS, Wells JM, Whittle JC, Whooley MA, Xu J, Yeh SS, Aslan M, Brewer JV, Brophy MT, Connor T, Argyres DP, Do NV, Hauser ER, Humphries DE, Selva LE, Shayan S, Stephens B, Whitbourne SB, Zhao H, Moser J, Beckham JC, Breeling JL, Romero JPC, Huang GD, Ramoni RB, Muralidhar S, Aguayo SM, Ahuja SK, Pyarajan S, Sun YV, Cho K, Gaziano JM, Wilson PW, O'Donnell CJ, Chang KM, Tsao PS, Assimes TL, Ritchie MD, Voight BF, The VA Million Veteran Program. Multi-trait association studies discover pleiotropic loci between Alzheimer's disease and cardiometabolic traits. Alzheimers Res Ther. 2021 Feb 4;13(1):34. PMCID: PMC7860582

19. Mirabnahrazam G, Ma D, Lee S, Popuri K, Lee H, Cao J, Wang L, Galvin JE, Beg MF, Initiative the ADN. Machine Learning Based Multimodal Neuroimaging Genomics Dementia Score for Predicting Future Conversion to Alzheimer’s Disease. J Alzheimers Dis. IOS Press; 2022 Jan 1;87(3):1345–1365. PMCID: PMC9195128

20. Meng L, Wang Z, Ji HF, Shen L. Causal association evaluation of diabetes with Alzheimer's disease and genetic analysis of antidiabetic drugs against Alzheimer's disease. Cell Biosci. 2022 Mar 10;12(1):28. PMCID: PMC8908591

21. Hu Z, Jiao R, Wang P, Zhu Y, Zhao J, De Jager P, Bennett DA, Jin L, Xiong M. Shared Causal Paths underlying Alzheimer's dementia and Type 2 Diabetes. Sci Rep. 2020 Mar 5;10(1):4107. PMCID: PMC7058072

22. Ma D, Lu D, Popuri K, Wang L, Beg MF, Alzheimer's Disease Neuroimaging Initiative. Differential Diagnosis of Frontotemporal Dementia, Alzheimer's Disease, and Normal Aging Using a Multi-Scale Multi-Type Feature Generative Adversarial Deep Neural Network on Structural Magnetic Resonance Images. Front Neurosci. 2020 Oct 22;14:853. PMCID: PMC7643018

23. Popuri K, Ma D, Wang L, Beg MF. Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. Hum Brain Mapp. 2020 Oct;41(14):4127–4147. PMCID: PMC7469784

24. Ma D, Popuri K, Bhalla M, Sangha O, Lu D, Cao J, Jacova C, Wang L, Beg MF, Alzheimer's Disease Neuroimaging Initiative. Quantitative assessment of field strength, total intracranial volume, sex, and age effects on the goodness of harmonization for volumetric analysis on the ADNI database. Hum Brain Mapp. 2019 Apr 1;40(5):1507–1527. PMCID: PMC6449147

25. Ma D, Yee E, Stocks JK, Jenkins LM, Popuri K, Chausse G, Wang L, Probst S, Beg MF. Blinded Clinical Evaluation for Dementia of Alzheimer's Type Classification Using FDG-PET: A Comparison Between Feature-Engineered and Non-Feature-Engineered Machine Learning Methods. J Alzheimers Dis JAD. 2021;80(2):715–726. PMCID: PMC8978589

26. Sangha O, Ma D, Popuri K, Stocks J, Wang L, Beg MF. Structural volume and cortical thickness differences between males and females in cognitively normal, cognitively impaired and Alzheimer's dementia population. Neurobiol Aging. 2021 Oct 1;106:1–11.

27. Ma D, Popuri K, Beg MF. Machine-learning-based longitudinal Dementia score based on brain atrophy pattern from structural MRI data. Mach-Learn-Based Longitud Dement Score Based Brain Atrophy Pattern Struct MRI Data.

28. Yee E, Ma D, Popuri K, Wang L, Beg MF, Initiative for the ADN, Ageing the AIB and L flagship study of. Construction of MRI-Based Alzheimer's Disease Score Based on Efficient 3D Convolutional Neural Network: Comprehensive Validation on 7,902 Images from a Multi-Center Dataset. J Alzheimers Dis. IOS Press; 2021 Jan 1;79(1):47–58. PMCID: PMC9159475

29. Ferreira D, Pereira JB, Volpe G, Westman E. Subtypes of Alzheimer’s Disease Display Distinct Network Abnormalities Extending Beyond Their Pattern of Brain Atrophy. Front Neurol [Internet]. 2019 [cited 2022 Jan 13];10. Available from: https://www.frontiersin.org/article/10.3389/fneur.2019.00524

30. Park JY, Na HK, Kim S, Kim H, Kim HJ, Seo SW, Na DL, Han CE, Seong JK. Robust Identification of Alzheimer's Disease subtypes based on cortical atrophy patterns. Sci Rep. 2017 Mar 9;7(1):43270.

31. Collij LE, Salvadó G, Wottschel V, Mastenbroek SE, Schoenmakers P, Heeman F, Aksman L, Wink AM, Berckel BNM, Flier WM van de, Scheltens P, Visser PJ, Barkhof F, Haller S, Gispert JD, Alves IL, Initiative for the ADN, Study for the A. Spatial-Temporal Patterns of β-Amyloid Accumulation: A Subtype and Stage Inference Model Analysis. Neurology. Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology; 2022 Apr 26;98(17):e1692–e1703. PMCID: PMC9071373

32. Vogel JW, Young AL, Oxtoby NP, Smith R, Ossenkoppele R, Strandberg OT, La Joie R, Aksman LM, Grothe MJ, Iturria-Medina Y, Pontecorvo MJ, Devous MD, Rabinovici GD, Alexander DC, Lyoo CH, Evans AC, Hansson O. Four distinct trajectories of tau deposition identified in Alzheimer's disease. Nat Med. 2021 May;27(5):871–881.

33. Cedres N, Ekman U, Poulakis K, Shams S, Cavallin L, Muehlboeck S, Granberg T, Wahlund LO, Ferreira D, Westman E, Initiative ADN. Brain Atrophy Subtypes and the ATN Classification Scheme in Alzheimer's Disease. Neurodegener Dis. Karger Publishers; 2020;20(4):153–164. PMID: 33789287

34. Ma D, Popuri K, Bhalla M, Sangha O, Lu D, Cao J, Jacova C, Wang L, Beg MF, Initiative ADN. Quantitative assessment of field strength, total intracranial volume, sex, and age effects on the goodness of harmonization for volumetric analysis on the ADNI database. Hum Brain Mapp. John Wiley & Sons, Inc. Hoboken, USA; 2019;40(5):1507–1527.

35. Ramchoun H, Ghanou Y, Ettaouil M, Janati Idrissi MA. Multilayer Perceptron: Architecture Optimization and Training. International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI); 2016 Sep [cited 2022 Dec 7]; Available from: https://reunir.unir.net/handle/123456789/11569

36. Crawshaw M. Multi-Task Learning with Deep Neural Networks: A Survey. ArXiv200909796 Cs Stat [Internet]. 2020 Sep 10 [cited 2022 Apr 27]; Available from: http://arxiv.org/abs/2009.09796

37. Young AL, Marinescu RV, Oxtoby NP, Bocchetta M, Yong K, Firth NC, Cash DM, Thomas DL, Dick KM, Cardoso J, van Swieten J, Borroni B, Galimberti D, Masellis M, Tartaglia MC, Rowe JB, Graff C, Tagliavini F, Frisoni GB, Laforce R, Finger E, de Mendonça A, Sorbi S, Warren JD, Crutch S, Fox NC, Ourselin S, Schott JM, Rohrer JD, Alexander DC. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. Nat Commun. 2018 Oct 15;9(1):4273.

38. Archetti D, Young AL, Oxtoby NP, Ferreira D, Mårtensson G, Westman E, Alexander DC, Frisoni GB, Redolfi A, for Alzheimer's Disease Neuroimaging Initiative and EuroPOND Consortium. Inter-Cohort Validation of SuStaIn Model for Alzheimer's Disease. Front Big Data [Internet]. 2021 [cited 2022 Feb 24];4. Available from: https://www.frontiersin.org/article/10.3389/fdata.2021.661110

39. Nobel L, Roblin DW, Becker ER, Druss BG, Joski PI, Allison JJ. Index of cardiometabolic health: a new method of measuring allostatic load using electronic health records. Biomarkers. Taylor & Francis; 2017 Jul 4;22(5):394–402. PMCID: PMC5676305

40. Ko W, Jung W, Jeon E, Mulyadi AW, Suk HI. ENGINE: Enhancing Neuroimaging and Genetic Information by Neural Embedding. 2021 IEEE Int Conf Data Min ICDM. 2021. p. 1162–1167.

41. Ko W, Jung W, Jeon E, Suk HI. A Deep Generative–Discriminative Learning for Multimodal Representation in Imaging Genetics. IEEE Trans Med Imaging. 2022 Sep;41(9):2348–2359.

42. Mirabnahrazam G, Ma D, Beaulac C, Lee S, Popuri K, Lee H, Cao J, Galvin JE, Wang L, Beg MF. Predicting time-to-conversion for dementia of Alzheimer's type using multi-modal deep survival analysis. Neurobiol Aging. 2023 Jan 1;121:139–156.

43. Silver M, Janousova E, Hua X, Thompson PM, Montana G. Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression. NeuroImage. 2012 Nov 15;63(3):1681–1694. PMCID: PMC3549495

44. Martinez-Murcia FJ, Ortiz A, Gorriz JM, Ramirez J, Castillo-Barnes D. Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders. IEEE J Biomed Health Inform. 2020 Jan;24(1):17–26.

45. Ausmees K, Nettelblad C. A deep learning framework for characterization of genotype data. G3 GenesGenomesGenetics. 2022 Mar 1;12(3):jkac020. PMCID: PMC8896001

46. Pirmoradi S, Teshnehlab M, Zarghami N, Sharifi A. A Self-organizing Deep Auto-Encoder approach for Classification of Complex Diseases using SNP Genomics Data. Appl Soft Comput. 2020 Dec 1;97:106718.

47. Song M, Greenbaum J, Luttrell J, Zhou W, Wu C, Luo Z, Qiu C, Zhao LJ, Su KJ, Tian Q, Shen H, Hong H, Gong P, Shi X, Deng HW, Zhang C. An autoencoder-based deep learning method for genotype imputation. Front Artif Intell [Internet]. 2022 [cited 2022 Dec 2];5. Available from: https://www.frontiersin.org/articles/10.3389/frai.2022.1028978

48. Dias R, Evans D, Chen SF, Chen KY, Loguercio S, Chan L, Torkamani A. Rapid, Reference-Free human genotype imputation with denoising autoencoders. Stephens M, Przeworski M, editors. eLife. eLife Sciences Publications, Ltd; 2022 Sep 23;11:e75600. PMCID: PMC9555874

49. Battey CJ, Coffing GC, Kern AD. Visualizing population structure with variational autoencoders. G3 GenesGenomesGenetics. 2021 Jan 1;11(1):jkaa036. PMCID: PMC8022710

50. Sundararajan M, Taly A, Yan Q. Axiomatic Attribution for Deep Networks [Internet]. arXiv; 2017 [cited 2022 Dec 7]. Available from: http://arxiv.org/abs/1703.01365

51. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis. 2020 Feb;128(2):336–359.