This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

Disentangling Neurodegeneration with Brain Age Gap Prediction Models

A Graph Signal Processing Perspective

Saurabh Sihag, Gonzalo Mateos, and Alejandro Ribeiro[†],[*]

Neurodegeneration is the progressive loss of structure or function of neurons in the brain. Reduction in cortical thickness or volume over time has been a workhorse metric to assess neurodegeneration in clinical settings; see also Case Study 1 for a demonstration of cortical atrophy assessment in the context of Alzheimer's disease (AD) relative to healthy individuals (HC group). Naturally, visual inspection of T1-weighted brain magnetic resonance imaging (MRI) images and associated MRI quantification products are used along with other biological measurements to make a 'subjective' assessment about the brain health of an individual. These assessments tend to be subjective because they lack a deterministic relationship between an individual's health status and the absolute values of metrics observed within MRI scans [1]. Moreover, such methods cannot adequately account for the statistical complexities inherent within neuroimaging datasets that capture neurodegeneration. In particular, neurodegeneration is a characteristic of the healthy aging process and various neurological disorders [2], exhibiting correlated patterns across brain regions. Such statistical factors motivate well the use of data-driven methods to characterize neurodegeneration.

Automating or improving the analyses of brain MRI images is appealing for several reasons: MRI is a non-invasive procedure and there is an untapped potential to reduce radiologists' missed detection error rates, leading to overall better patient treatment and outcomes, to name a few. In this article, we focus on the family of 'Brain Age Gap Prediction' models. In simple terms, these models use machine learning (ML) algorithms to process neuroimaging data with the goal of predicting how much older the brain of an individual is relative to their chronological age; the difference being the so-termed brain age gap. Brain age gap prediction models have recently gained traction in digital health and personalized medicine, due to their promise of generating informative, yet compact, summary statistics of brain health for clinical use. Specifically, these models are hypothesized to leverage anomalous patterns associated with neurodegeneration in imaging data to yield a biomarker representative of the extent of neurodegeneration within an individual. This hypothesis has been corroborated in multiple recent studies, where the brain age gap (also, sometimes referred to as brain age delta) has been shown to be predictive of disease severity.

[†]Saurabh Sihag is with the Department of Electrical and Computer Engineering at the University at Albany, SUNY, Albany, NY (email: ssihag@albany.edu). Gonzalo Mateos is with the Department of Electrical and Computer Engineering at the University of Rochester, Rochester, NY (email:gmateosb@ece.rochester.edu). Alejandro Ribeiro is with the Department of Electrical and Systems Engineering at the University of Pennsylvania, Philadelphia, PA (email:aribeiro@seas.upenn.edu).

[*] for the Alzheimer's Disease Neuroimaging Initiative (see Acknowledgement section).

To offer the required background and context, this tutorial begins with an overview of the brain age gap prediction algorithm and a survey of various existing studies that reinforce its relevance to characterizing neurodegeneration; see 'Brain Age Gap Prediction Models'.

Despite wide-ranging promising results, there are several challenges facing the practical deployment and generalizability of brain age gap prediction models in clinically meaningful settings (e.g., heterogeneous populations with distinct health conditions). Notable are methodological obscurities driven by the lack of a deterministic relationship (or conceptual justification) that ties the accuracy of the ML model for age prediction, to its usefulness in deriving a clinically significant brain age gap in neurodegeneration. In this context, we will review the relevant evidence from the literature and argue that such roadblocks to the practical adoption of brain age gap prediction stem from the opaqueness of the ML models used. The main contribution of this tutorial is to identify key mathematical principles that can help overcome the aforementioned challenges by bringing to bear graph signal processing (GSP).

Recent advances in GSP have introduced a breadth of *principled* analytical tools for graph-structured (or correlated multivariate) data, which match well with the intricacies of neuroimaging datasets. For instance, the cortical thickness features in Case Study 1 can be interpreted as *graph signals* over some graph where nodes represent cortical brain regions (see Fig. 5). Edges are defined using the pairwise correlations between regional anatomical features, i.e., the entries of the anatomical covariance matrix that will be affected by brain atrophy. We contend GSP offers a natural framework to study anatomical brain features and to bridge key methodological gaps in brain age gap prediction. To this end, deep learning methods that build on GSP foundations take center stage. In 'GSP Foundations for Neuroimaging Data Analysis' we review graph neural networks (GNNs) with convolutional layers and survey their theoretical properties, positioning them as an attractive tool for neuroimaging data analysis. We ground these discussions by introducing a GNN that is instantiated on an anatomical covariance matrix, called coVariance neural network (VNN). This way, a covariance matrix estimated from anatomical features derived from structural MRI is leveraged as a graph representation of signal structure. Crucially, we discuss the rich theoretical properties of VNNs and how they translate to unique operational capabilities that facilitate transparent and robust construction of brain age gap in neurodegeneration. Our discussions converge to a detailed overview of an explanation-driven brain age gap prediction pipeline; see 'Towards Explainable Brain Age Gap Prediction from Structural MRI'. The exposition is supported by various case studies to demonstrate how key methodological obscurities in this application domain are overcome using VNN models.

All in all, this tutorial article elucidates the intellectual depth and clarifications added to brain age gap prediction algorithms via an interdisciplinary perspective rooted at the crossroads of GSP, ML theory, and network neuroscience. We conclude with brief discussions on adopting GSP for future studies in frontier applications such as domain-specific foundation models and neurodegenerative disease subtyping.

Case Study 1: Cortical atrophy characterizes neurodegeneration in Alzheimer's disease (AD)
In this case study, we leverage the dataset from the ADNI study [3] to demonstrate cortical atrophy in AD. This dataset consisted of three cohorts: (a) 206 healthy individuals (HC; age =

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

73.87 ± 6.39 years, 110 females); (b) 372 individuals diagnosed with mild cognitive impairment (MCI; age = 72.26 ± 7.61 years, 160 females); and (c) 118 individuals diagnosed with AD (age = 73.84 ± 7.56 years, 56 females). MCI diagnosis represents an early stage of loss of cognitive ability and is a precursor to AD. For each individual, 68 cortical thickness features were available. These features are publicly available at https://adni.loni.usc.edu/ and had been derived by processing T1-weighted structural MRI scans via Freesurfer software [4]. The cortical thickness features were curated according to the Desikan-Killiany brain atlas [5].

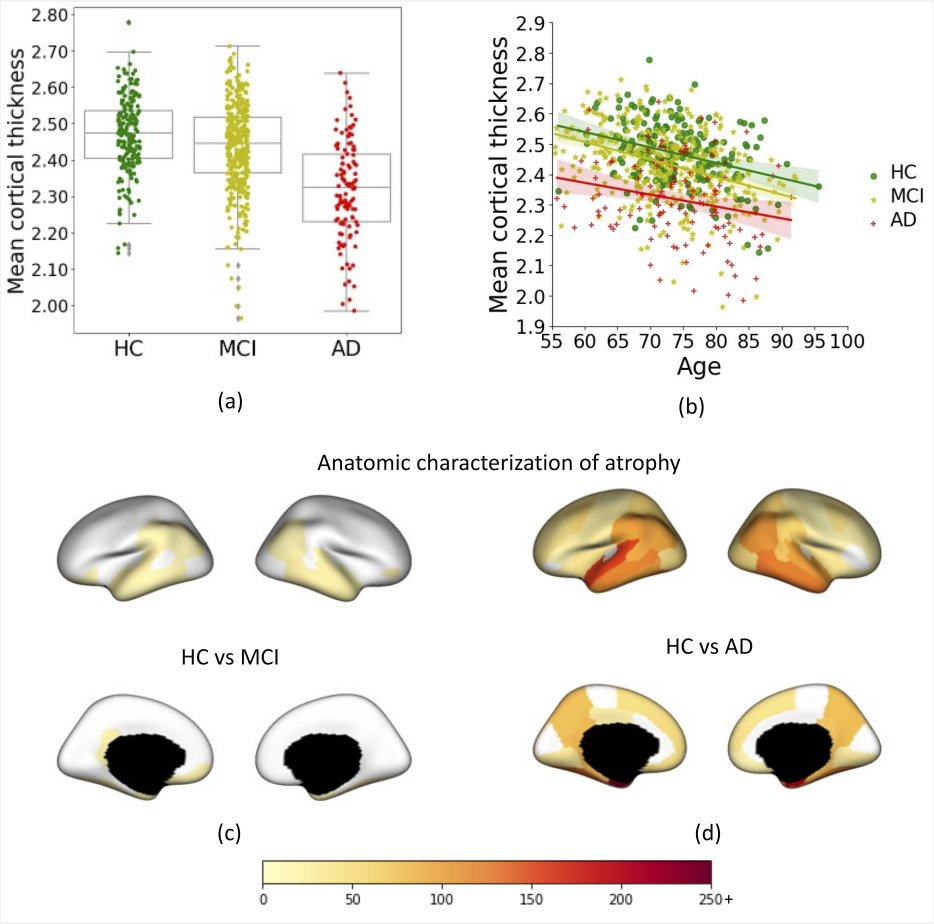

Fig. 1. (a) Distributions of mean cortical thickness across the 68 cortical regions in HC, MCI, and AD cohorts. (b) Scatterplot between the mean cortical thickness and age for the HC, MCI, and AD cohorts. The lines representing the linear fit for individual cohorts are also shown. (c) Brain atrophy as derived from cortical thickness in MCI cohort relative to HC cohort. (d) Brain atrophy as derived from cortical thickness in AD cohort relative to HC cohort. In (c) and (d), the F-values associated with statistically significant group differences in cortical thickness between MCI or AD groups and HC group as given by ANCOVA with age as covariate (p-value after Bonferroni correction < 0.05) have been projected on the brain surface.

Figure 1 summarizes the characterization of neurodegeneration via brain atrophy as determined by cortical thickness features. Figure 1a illustrates the distributions of mean cortical metrics (across the whole brain) for the HC, MCI, and AD cohorts. With the reduction in mean cortical thickness

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

representing brain atrophy, it is apparent that the AD group exhibited higher brain atrophy than the HC group, with the MCI group falling in between them. Moreover, mean cortical thickness metrics for all groups exhibited negative linear relationships with age (Fig. 1b), suggesting that neurodegeneration was a characteristic of aging across all groups. Figures 1c and 1d provide the anatomic characterizations of brain atrophy in terms of cortical thickness features in MCI and AD groups. The MCI group exhibited statistically significant reduction in cortical thickness relative to HC group (ANCOVA with age as covariate, p-value after Bonferroni correction < 0.05) in bilateral medial temporal lobe and temporo-parietal junction regions. Similar analysis revealed more prominent brain atrophy across a majority of brain regions in AD relative to HC in Fig. 1d, with the most prominent regions of atrophy including the bilateral entorhinal and medial temporal lobe. The contrast in the effect sizes of brain atrophy for MCI group in Fig. 1c and AD group in Fig. 1d is reasonable as the MCI diagnosis is typically a precursor of AD diagnosis.

BRAIN AGE GAP PREDICTION MODELS

Data from various neuroimaging modalities, including structural MRI, functional MRI, and positron emission tomography (PET), are able to capture the changes within various facets of the brain due to neurodegeneration and healthy aging [6]. We henceforth focus on structural MRI since it provides high-quality anatomical details of the brain and is among the most widely adopted modalities in clinical workflows. It also has potential diagnostic utility as features derived from structural MRI (brain atrophy, for example) can differentiate neurodegenerative conditions from healthy aging [7]. The baseline for healthy aging is provided by the progressive anatomical and functional changes in the brain captured by neuroimaging datasets over the lifespan [8].

Within the landscape of data-driven ML algorithms that use structural MRI to identify neurodegeneration biomarkers, brain age gap prediction specifically targets the hypothesis that individuals can age *biologically* at variable rates [9].

Neurodegeneration markers within structural MRI can be linked to the phenomenon of *accelerated aging*, i.e., when the brain imaging data of an individual reflect patterns consistent with an advanced age relative to their current chronological age. For instance, brain atrophy is characterized by loss of cortical thickness and volume metrics, a characteristic of both healthy aging and neurodegeneration. Accelerated brain atrophy, often concentrated in specific brain regions, is a distinguishing feature of neurodegeneration relative to healthy aging. Hence, regions of the brain with accelerated atrophy relative to healthy individuals can be (statistically) perceived to have experienced accelerated aging. This phenomenon is apparent in Fig. 1, where we observe a lower mean cortical thickness in the AD cohort relative to the HC group (Fig. 1a), with the reduction in cortical thickness concentrated in certain areas (Fig. 1d). In general, accelerated aging has been recognized as a predictor of morbidity and impairment [10].

Brain age gap prediction algorithms provide viable methods to quantify accelerated aging. Specifically, they generate a compact scalar-valued representation of the biological age of an individual from features

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

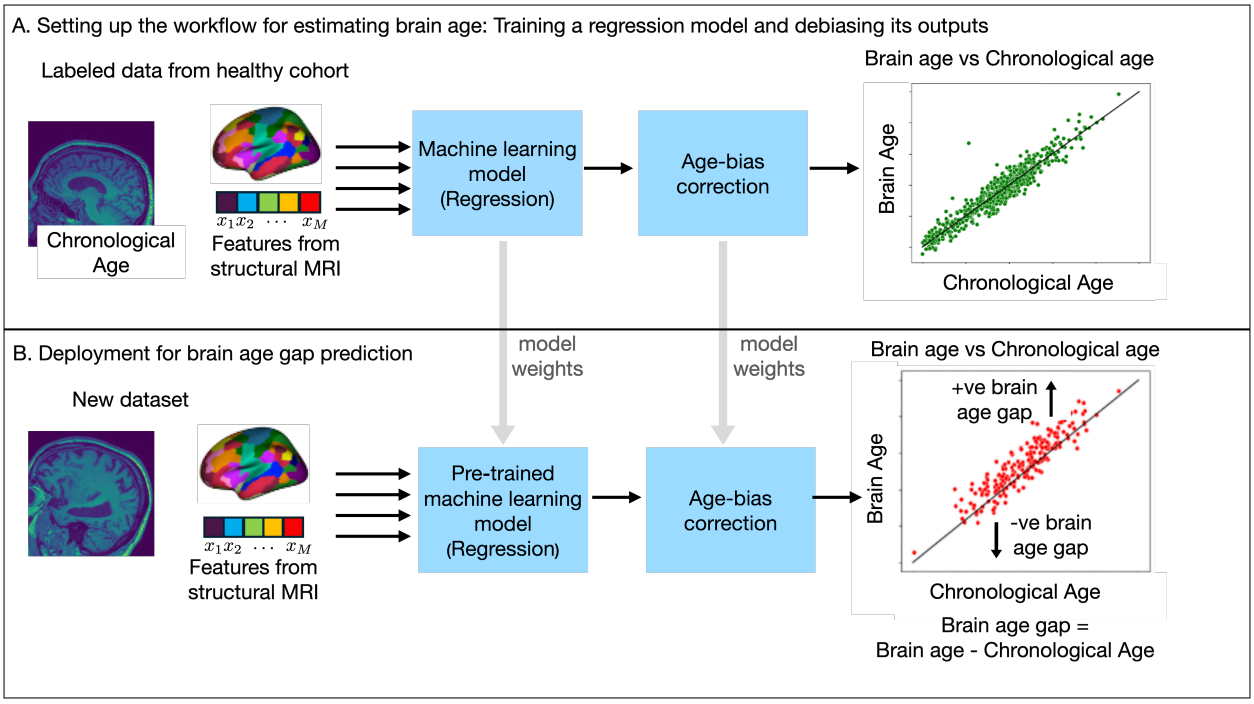

Fig. 2. Schematic of brain age gap prediction using ML. (A) Neuroimaging data, formed by T1-weighted structural MRI scans, from a set of healthy individuals are labelled with their respective chronological age. Pre-processing pipelines using standard tools, such as Freesurfer, may be applied to extract relevant features from the MRI scan. A regression model is trained using the extracted features or the raw MRI scans, as preferred. The outputs of the ML model are then corrected for any age biases using an appropriate statistical correction procedure. Note that the age-bias correction is applied after training the regression model. The age-bias corrected outputs form the estimate of the brain age. (B) The trained ML model and its associated age-bias correction module can then be deployed to predict brain age using neuroimaging data pre-processed from a new dataset. The brain age gap is obtained as the difference between the predicted brain age and the chronological age.

derived from structural MRI [11]. The metric of interest is the 'Brain Age Gap', given by

$$\text{Brain Age Gap} \triangleq \text{Predicted Brain Age} - \text{Chronological Age}, \tag{1}$$

where predicted brain age is the estimate of biological age derived from neuroimaging data [12]. In this context, brain age gap prediction is also often referred to as 'brain age prediction' or 'brain clock prediction', as the chronological age of an individual is usually known. Naturally, brain age gap prediction algorithms aim to extract traces of biological aging inherent in MRI scans.

How is brain age gap evaluated?

Various facets of the schematic brain age gap prediction workflow in Fig. 2 are highlighted next. This workflow is primarily motivated by the hypothesis that an ML model pre-trained to gauge healthy aging can detect accelerated aging (i.e., infer brain age $>$ chronological age).

Data curation. The training set for the ML model consists of the chronological age and the features derived from the structural MRI scans of a cohort of healthy individuals. The MRI scans may be pre-processed via image processing pipelines (such as Freesurfer [4] and CAT12 [13]) to

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

extract meaningful features predictive of aging (for example, brain volume or thickness at each voxel). Moreover, the extracted features may be organized anatomically according to a pre-selected brain atlas [5]. Some ML pipelines directly operated on the raw MRI scans [14].

Training the ML model as a regression model. The features extracted from structural MRI are used as predictors in a regression model trained to predict the chronological age of the healthy population. This pre-trained model provides an estimate $\hat{y}$ for an individual with chronological age y. The regression model is selected from the class of ML approaches suitable for multivariate data analyses, such as support vector regression, principal component analysis (PCA)-based regression, GNNs, or convolutional neural networks (CNNs). The loss function penalizes the (e.g., mean-squared error) deviation between the predicted outcome $\hat{y}$ and the chronological age y.

Age-bias correction. The predictions generated by the regression model for the healthy population are further evaluated for *age bias*, which may arise when the correlation between predicted age and chronological age is markedly smaller than 1. In this scenario, the age of younger individuals may be overestimated while those of older individuals may be underestimated. This statistical bias correction could readily be applied with an appropriate linear model [15]. A typical two-step age bias correction procedure to obtain the brain age prediction $\hat{y}_{\mathsf{B}}$ operates as follows [15]:

Step 1. Fit a linear model to the training set to estimate scalars ω and ϱ in: $\hat{y} - y \sim \omega y + \varrho$.

Step 2. Evaluate brain age $\hat{y}_{\mathsf{B}}$ for an individual with chronological age y and chronological age estimate $\hat{y}$ from their structural MRI features as follows:

$$\hat{y}_{\mathsf{B}} = \hat{y} - (\omega y + \varrho) \; . \tag{2}$$

The difference between $\hat{y}_{\mathsf{B}}$ and y is the brain age gap, henceforth denoted as Δ-Age, i.e.,

$$\Delta\text{-Age} \triangleq \hat{y}_{\mathsf{B}} - y. \tag{3}$$

Deployment of the brain age gap prediction pipeline. Features derived from the structural MRI scans of a new set of individuals (possibly with an unknown health status) can now be fed to the regression model with appropriate age-bias correction to generate the individual predictions of brain age. The difference between predicted brain age and chronological age quantifies the brain age gap. For example, if the predicted brain age of a 60-years-old individual is 70 years, we say the brain age gap or Δ-Age for this individual is $+10$ years.

Brain age gap as a biomarker of neurodegeneration

When trained on a healthy population, the ML model is expected to learn the statistical patterns of healthy aging. Using the learned representations of healthy aging as a reference point, the inferred Δ-Age compactly captures accelerated aging if it has a smaller magnitude in a healthy population, and drifts toward a larger magnitude in the specific direction of $\Delta\text{-Age} > 0$ for individuals with neurodegeneration [12].

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

The validity of brain age gap (Δ-Age) as a biomarker of neurodegeneration hinges on its characterization of clinical markers of disease severity or risk and underlying biological processes.

The usefulness of Δ-Age as a biomarker is often justified by demonstrating a combination of (i) higher Δ-Age in neurodegeneration relative to the healthy population; and (ii) characterizing the clinical or biological markers of disease burden by post-hoc analyses of Δ-Age; see Table I for a summary of such representative works. The former provides statistical evidence of accelerated aging in neurodegeneration, and (ii) is essential for interpretability. These aspects are illustrated in Case Study 2.

Case Study 2: Interpreting Δ-Age as a biomarker in AD

ML model. We consider GNN-based regression for brain age gap prediction from cortical thickness features in the ADNI dataset; see Case Study 1 and [16]. A GNN is chosen to exploit the correlation-induced information structure among the cortical-thickness features. The model was trained to predict the chronological age of the cognitively healthy population from the OASIS-3 dataset [17] (611 individuals, age = 68.38 ± 7.62 years, 351 females), using their cortical thickness features curated according to the Desikan-Killiany atlas.

Δ-Age in AD. The pre-trained model was used to predict Δ-Age for different cohorts in the ADNI dataset described in Case Study 1 [3]. We also investigated the relationship between Δ-Age and Clinical dementia rating- sum of boxes (CDRSB) metrics to assess the clinical interpretation of Δ-Age. CDRSB is a clinical marker to stage dementia severity [18]. CDRSB for MCI (available for 200 individuals, mean = 1.33, standard deviation = 0.94) was substantially smaller relative to that for the AD cohort (available for 70 individuals, mean = 5.61, standard deviation = 2.35).

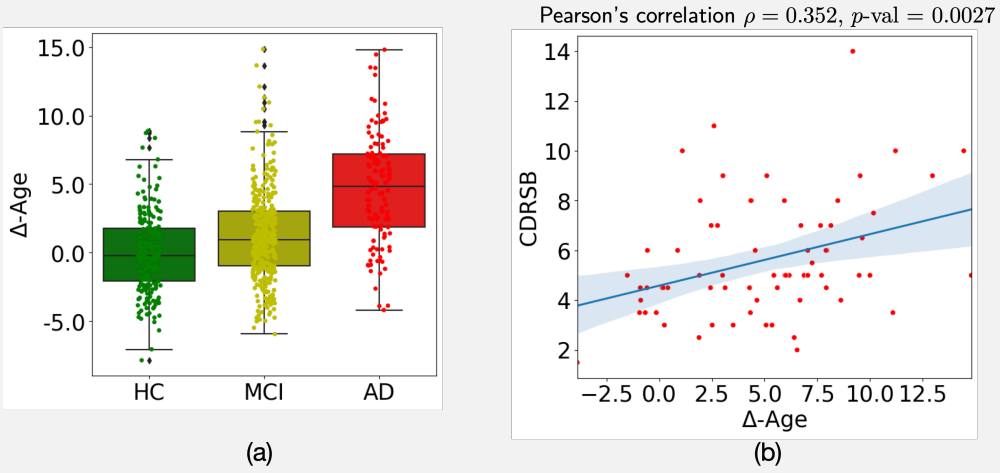

Fig. 3. (a) Distributions of Δ-Age in HC, MCI, and AD cohorts. (b) Scatterplot between the CDRSB and Δ-Age for the AD cohort. The line representing the linear fit is also shown.

Figure 3 reproduces the results from [16] and validates that Δ-Age inferred by the GNN model is a biomarker of AD. In fact, the Δ-Age distribution has the highest mean for the AD group (Δ-Age = 4.66 ± 4.04 years) and the smallest mean for the HC group (Δ-Age = 0 ± 2.9 years);

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

the MCI cohort lies in between them (Δ-Age = 1.23 ± 3.29 years). Furthermore, Δ-Age in the AD group was significantly correlated with CDRSB (Pearson's correlation = 0.352, p-val = 0.0027).

The studies in [19]–[21] primarily considered AD as the disease group and consistently reported elevated Δ-Age in AD relative to the healthy population. Moreover, the study in [20] stratified the population according to APOE ε_4 carriers and non-carriers (a gene variant that increases the risk of developing AD) to report varying Δ-Age in these subgroups, which also exhibited different patterns of longitudinal progression of the disease. The focus of [21] was to demonstrate the anatomical interpretability of brain age in AD and different subgroups within the healthy population, through prevalent methods from explainable artificial intelligence (AI). Similar analyses were adopted in other disease-specific Δ-Age studies for schizophrenia [22], TBI [23], and Parkinson's disease [24].

As pre-trained models, brain age gap prediction algorithms are widely applicable to derive biomarkers for numerous clinically-defined health conditions associated with neurodegeneration.

A common theme in all the aforementioned studies is that the brain age gap prediction model was trained only on a healthy population, and then deployed to study Δ-Age for specific neurodegenerative conditions. Hence, the ML workflow for Δ-Age prediction has a generalist characteristic as it is not bound to a specific health condition, unlike many supervised learning algorithms developed to study neurodegeneration. This key observation has been well documented [21], [24], [25], as multiple studies derived anatomical patterns to confirm that disease-relevant features contributed to the reported brain age (or Δ-Age). Higher Δ-Age was found for AD, FTD, and LBD in [24], followed by distinct anatomical characterizations of these neurodegenerative conditions using explainable AI tools. In a similar spirit, the study in [25] leveraged a GSP-driven explainable model for Δ-Age prediction [26] to report distinct anatomical characterizations of Δ-Age in AD, FTD, and the combined group of CBS and PSP conditions. Existing studies have also shown the effectiveness of brain age gap in other clinical domains, such as for major depressive disorder [27] and to track changes throughout the human lifespan [28]. For a comprehensive review of brain age gap prediction algorithms applied to different biological domains, the interested reader is referred to [29]. Moreover, brain age gap has been discussed within the broader context of biological age estimation [30]. In summary, the body of work surveyed in this section justifies the adoption of brain age gap as a promising biomarker for the early stages of clinical workflows, where individuals may not yet have a clear diagnosis.

Performance-driven approach to brain age gap prediction: Unresolved challenges and gaps

Despite promising results in characterizing neurodegeneration, there exists considerable divergence in the underlying ML principles adopted for Δ-Age prediction [32]. The diversity of ML models chosen for this application notwithstanding (see e.g., [33] for a recent review), here we elucidate the fundamental methodological obscurities in Δ-Age prediction by performance-driven principles that severely challenge their adoption in practice. These limitations have been well documented in the literature [26], [34], [35].

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

TABLE I

A SUMMARY OF THE STUDIES ON Δ-AGE FOR VARIOUS NEURODEGENERATIVE CONDITIONS AND ASSOCIATED EXPERIMENTS THAT VALIDATE Δ-AGE AS A BIOMARKER.

Clinical Condition	**Experiments establishing Δ-Age as a Biomarker**	**Reference**
AD	Elevated Δ-Age in disease group	[19]
AD	Differentiating Δ-Age in APOE ε_4 carriers/non-carriers, longitudinal progression, correlation with neuropsychological test scores	[20]
AD	Elevated Δ-Age in disease group, anatomical maps, associations with neurocognitive measures	[21]
AD, Frontotemporal Dementia (FTD), Lewy Body Dementia (LBD)	Elevated Δ-Age in disease groups, distinct anatomical maps for Δ-Age, associations with neurocognitive measures, tau PET	[31]
Schizophrenia	Elevated Δ-Age in disease groups, longitudinal analysis	[22]
Traumatic Brain Injury (TBI)	Elevated Δ-Age in TBI, correlation with time since injury	[23]
Parkinson's Disease (PD)	Elevated Δ-Age in disease group, associations with cognitive and motor impairment	[24]
AD, FTD, Corticobasal Syndrome (CBS), Progressive Supranuclear Palsy (PSP)	Elevated Δ-Age in disease groups, distinct anatomical maps for Δ-Age in disease groups	[25]

In light of the challenge stemming from the lack of a tangible ground truth for brain age, the focus of most ML-driven approaches in this domain has overwhelmingly been on: (i) achieving near perfect performance on the chronological age prediction task for healthy individuals; and (ii) using it as a measure to gauge the quality of a Δ-Age prediction framework. Here, we use the taxonomy 'performance-driven approach' to Δ-Age prediction for such ML methods. Performance-driven approaches gauge their quality through metrics such as mean absolute error (MAE) on chronological age prediction in healthy populations. They encompass both traditional ML-driven methods and deep learning models, and the higher expressive power of the latter makes them the prevalent choice today [33].

In principle, achieving the best possible performance in predicting chronological age in a healthy population is well motivated, as it allows the ML model to learn patterns of healthy aging. However, as is abundantly clear from Table I and our prior discussion in 'Brain age gap as a biomarker of neurodegeneration', the *validation* of Δ-Age as a biomarker of health conditions is critical for its meaningful adoption in clinical settings.

For Δ-Age to be a valid biomarker, it is unclear how accurate the underlying ML brain age gap prediction algorithm must be when used to predict chronological age for a healthy population.

There exists ample evidence in the literature to corroborate that a more accurate prediction of chronological age in healthy populations does not necessarily translate into improved validation of the inferred Δ-Age as a biomarker. The study in [34] reported that a Δ-age prediction model with a relatively 'moderate' fit on the chronological age of healthy individuals led to inferred Δ-Age with better clinical utility in neurodegenerative conditions, when compared to a model that had a 'tighter' fit on the chronological age

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

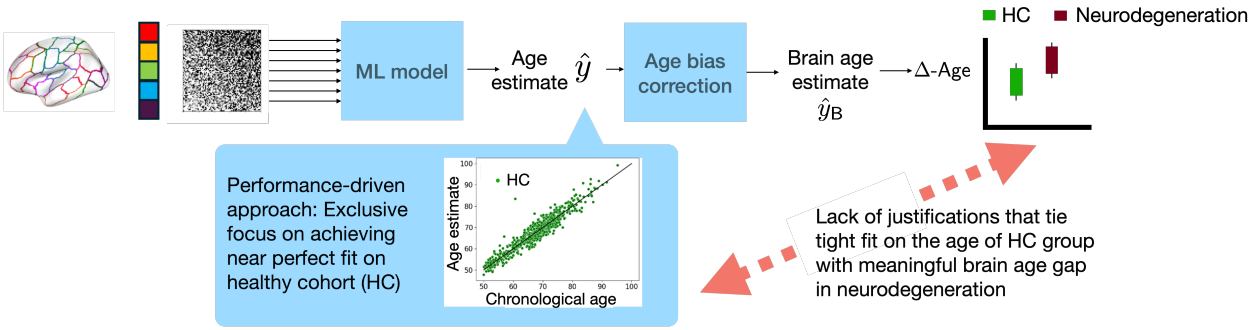

Fig. 4. Performance-driven approaches to Δ-Age prediction prioritize a near-perfect fit on the chronological age of the HC, yet they lack the conceptual or statistical justifications to ensure the relevance of inferred Δ-Age as a biomarker for neurodegeneration.

of the same healthy cohort. Similar findings were reported in [26], where the model with higher MAE on chronological age predictions of the healthy population inferred Δ-Age with a higher correlation with CDRSB scores in AD (relative to a baseline model with smaller MAE). A comprehensive evaluation of various ML approaches in [35] found no significant correlation between the accuracy of chronological age prediction and the clinical utility of the accompanying Δ-Age estimates. Figure 4 summarizes the methodological obscurity in Δ-Age prediction by performance-driven approaches.

Chronological age clearly acts as a proxy for biological age in healthy individuals when pre-training an ML model for Δ-Age prediction. However, the discussion above *does not* lead to the conclusion that pre-training the model on a healthy cohort for chronological age prediction is inherently flawed. A more nuanced observation to make is that the overwhelming focus on achieving a near-perfect fit on the chronological age of the healthy population potentially overlooks meaningful biological information necessary to discriminate between the healthy and clinical groups [32]. We contend that this could be attributable to performance-driven approaches' neglect of the heterogeneity of biological aging within the healthy population itself, due to factors such as obesity, stress levels, among others. These variables also affect the propensity to develop a neurodegenerative condition in the future, within the healthy population.

> Performance-driven approaches tend to take the leap from achieving the 'best possible fit' with a principled ML model to a 'near perfect fit' on chronological age by arbitrary additions of sophisticated computation modules, compounding methodological obscurities in Δ-Age prediction.

Recent performance-driven studies have adopted model-agnostic, post-hoc methods from explainable AI to corroborate the biological validity of their brain age predictions. For instance, saliency maps were utilized in [21] and [31] to identify those brain regions most relevant to predicted brain age for different clinical groups. However, such approaches still provide an incomplete perspective to Δ-Age as a biomarker. Specifically, since Δ-Age is a deviation from the chronological age, an explanation of brain age alone overlooks the component of the deviations due to healthy brain aging. Additionally, the explanations offered by post-hoc explainability methods suffer from lack of robustness, for example, due

to instability to small perturbations in the input, variability in explanations due to stochasticity in training algorithms, and model multiplicity (i.e., when multiple models with similar performance may exist but offer distinct explanations) [36]–[38]. Therefore, further progress is needed within the explainable AI domain for these approaches to be confidently adopted in practice.

Adding mathematical depth to brain age gap prediction. It is unlikely that the conceptual gap in performance-driven approaches regarding the statistical dependency between the accuracy of the model during training on healthy individuals and Δ-Age as a biomarker for neurodegeneration can be bridged with experiments alone. Therefore, the development of relevant mathematical principles is critical for a viable and generalizable practical methodology. To this end, we identify the following four mathematical principles:

- **Principle 1: Focusing on Δ-Age as a residual of regression tasks.** The residuals of the regression models inform the Δ-Age estimates when the ML model is deployed to predict chronological age for individuals with adverse health conditions. Hence, it is paramount to focus on the structure and statistics of the residuals of the age prediction model, rather than the age prediction task itself, to validate the viability of Δ-Age as a biomarker.
- **Principle 2: Shift focus to qualitative evaluation of ML models trained on the healthy population.** It is key to go beyond the performance on the chronological age prediction task and instead focus on a holistic characterization of the representations that the ML model learns when exposed to data from the healthy population.
- **Principle 3: Transference of pre-trained age prediction models to neurodegenerative cohorts for constructing Δ-Age as a biomarker.** In principle, predicting Δ-Age in neurodegenerative cohorts can be perceived as an unsupervised transfer learning problem, where we expect a degradation in performance. This problem is unsupervised because the expected amount of drift in model performance is unknown (i.e., there is no ground truth for brain age in neurodegenerative cohorts). It would be desirable to identify the specific deviations in the information processing pipeline itself, which are the contributors to the elevated Δ-Age observed in neurodegeneration.
- **Principle 4: Generalizability beyond specific dimensionality of the data.** Due to the existence of different brain atlases, the neuroimaging datasets characterizing the same population can have arbitrary dimensionalities in independently conducted studies [39]. This creates a challenge for reproducibility of findings, as the prevalent performance-driven approaches in both traditional ML and deep learning are limited within the dimensionality of the dataset that they have been trained on. To address this challenge, we need mathematical principles that govern the reproducibility of findings derived using an ML model across datasets of different dimensionalities, without being re-trained from scratch.

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

Adoption of Principles 1-4 in brain age gap prediction will fundamentally shift the primary focus to Δ-Age prediction as a biomarker, rather than being a byproduct of age prediction in performance-driven approaches.

Our discussion on 'Adding mathematical depth to brain age gap prediction' summarized desirable mathematical principles behind brain age gap prediction to improve its practical viability. Traditional ML methods or the prevalent deep learning models adhere to some, but not all the principles identified above. For instance, a PCA-regression model could address the requirements of Principle 2 via transparent evaluation of Δ-Age, but at the same time, such a model may suffer from instability [40]. On the other end of the spectrum, deep learning models can offer improved robustness and advanced operational capabilities, thus meeting the requirements of Principle 4. However, these 'black-box' architectures output predictions that are not inherently explainable; thus, they fall short in meeting the requirements of Principles 1-3. For instance, CNNs are commonly adopted in the brain age gap prediction pipeline [21], [34], but fail to convincingly reveal the anatomical factors contributing to brain age gap in neurodegeneration. Based on the preceding discussion, we argue that the alignment of Principles 1-4 with the brain age gap prediction pipeline is nontrivial and requires a deeper *theoretical* understanding of the chosen ML model. GSP offers an ideal suite of foundational tools for structured multivariate information processing and ML over graphs, facilitating the desired explainability, robustness, and transferability analyses.

GSP Foundations for Neuroimaging Data Analysis

Recent GSP advances have led to principled and theoretically sound learning tools for a variety of applications where data reside in non-Euclidean domains and exhibit graph structure [41]. GSP paved the way for innovative GNN architectures, thus, bridging signal processing insights and mathematical theory with the empirical successes of deep learning [42]. The domain of network neuroscience, which studies the brain via its network representations, has been a major beneficiary of these advances in GSP due to a concurrent increase in the availability of large spatiotemporal MRI datasets [43], [44].

GSP offers a natural substrate over which the mathematical advancements needed for a practically viable brain age gap prediction workflow can be developed.

GSP and GNNs: An overview

We review the background on GSP and GNNs needed to introduce an explainable brain age gap prediction framework adhering to Principles 1-4; see also [41], [42] for other insightful tutorial treatments.

The standard information processing backbone in GSP can be described in terms of four main pillars. First, consider a *graph* $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$ with a set of M nodes $\mathcal{V} = \{1, \ldots, M\}$, a set of undirected edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and an edge weight function $\mathcal{W} : \mathcal{E} \mapsto \mathbb{R}$. The graph topology can be compactly represented using a symmetric matrix $\mathbf{A}$ of size $M \times M$, which encodes the edge and weight information. The adjacency and Laplacian matrices are examples of such commonly used matrix representations. In neuroimaging data analysis, the graph is typically data-driven, with its nodes representing different brain regions and

the edges and weights inferred from nodal data; e.g., pairwise correlations among cortical thickness features (Fig. 5). Second, a *graph signal* $\mathbf{x} = [x_1, \ldots, x_M]^\top \in \mathbb{R}^M$ is a vector representation of the data supported on the graph $\mathcal{G}$, i.e., each element within $\mathbf{x}$ can be associated with a distinct node in $\mathcal{V}$. Going back to Case Study 1, the vector of cortical thickness features for an individual represents their graph signal. Third, a *graph filter* is the computational module to transform the graph signal $\mathbf{x}$ over the graph representation $\mathbf{A}$ [45]. Information processing with a graph filter relies on the *shift* operation $\mathbf{A}\mathbf{x}$, which shifts the graph signal $\mathbf{x}$ over the nodes in $\mathbf{A}$, such that, the i-th component of $\mathbf{A}\mathbf{x}$ is

$$[\mathbf{A}\mathbf{x}]_i = \sum_{j=1}^{M} [\mathbf{A}]_{ij} \mathbf{x}_j, \tag{4}$$

i.e., its value is determined by an aggregation of the information in $\mathbf{x}$ according to the weights in the i-th row of $\mathbf{A}$ (corresponding to edges incident to node i). In general, the output of the shift operation, $\mathbf{A}\mathbf{x}$, is another graph signal, whose elements are obtained by linear mixing of the elements in $\mathbf{x}$ according to the weights in $\mathbf{A}$. Building upon this observation, the graph filter implements the convolution operation via a polynomial on $\mathbf{A}$, such that, the output of the graph filter is

$$\mathbf{z} = \mathbf{H}(\mathbf{A})\mathbf{x}, \quad \text{where } \mathbf{H}(\mathbf{A}) = \sum_{k=0}^{K} h_k \mathbf{A}^k, \tag{5}$$

and $\mathbf{h} = [h_0, h_1, \ldots, h_K]^\top \in \mathbb{R}^K$ are the *filter taps*. Finally, the *graph Fourier transform* (GFT) facilitates a spectral decomposition of graph signals and filters. Consider the eigendecomposition $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$, where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_M]$ is the orthonormal matrix of M eigenvectors, and $\mathbf{\Lambda} = \mathsf{diag}(\lambda_1, \ldots, \lambda_M)$ is the diagonal matrix of eigenvalues ordered as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$. The graph Fourier transform (GFT) is defined as the projection of the signal $\mathbf{x}$ onto the eigenspace of $\mathbf{A}$, namely

$$\tilde{\mathbf{x}} = \mathbf{U}^\top \mathbf{x} \, . \tag{6}$$

The i-th entry of $\tilde{\mathbf{x}}$ quantifies the contribution of $\mathbf{u}_i$ to the graph signal $\mathbf{x}$ via the inner product $[\tilde{\mathbf{x}}]_i = \mathbf{u}_i^\top \mathbf{x}$. Indeed, from (6) it follows that a graph signal can be synthesized as $\mathbf{x} = \sum_{i=1}^{M} [\tilde{\mathbf{x}}]_i \mathbf{u}_i$. Eigenvector $\mathbf{u}_i$ is the frequency component associated to frequency λ_i, and $[\tilde{\mathbf{x}}]_i$ the corresponding GFT coefficient of $\mathbf{x}$.

Taking the GFT of the graph filter output in (5) and considering the eigendecomposition of $\mathbf{A}$ yields

$$\tilde{\mathbf{z}} = \mathbf{U}^\top \mathbf{z} = \mathbf{U}^\top \sum_{k=0}^{K} h_k \mathbf{U}\mathbf{\Lambda}^k \mathbf{U}^\top \mathbf{x} = \sum_{k=0}^{K} h_k \mathbf{\Lambda}^k \mathbf{U}^\top \mathbf{x} = \mathbf{H}(\mathbf{\Lambda})\tilde{\mathbf{x}}, \tag{7}$$

where $\mathbf{H}(\mathbf{\Lambda}) = \mathsf{diag}(h(\lambda_1), \ldots, h(\lambda_M))$ is a diagonal matrix and

$$h(\lambda_i) := \sum_{k=0}^{K} h_k \lambda_i^k \, . \tag{8}$$

By inspection of (6)-(8), it follows that the impact of the graph filter on the i-th element of $\tilde{\mathbf{x}}$ (the inner product $[\tilde{\mathbf{x}}]_i = \mathbf{u}_i^\top \mathbf{x}$) is limited to a scaling by $h(\lambda_i)$, in a way akin to a convolution theorem, i.e.,

$$[\tilde{\mathbf{z}}]_i = h(\lambda_i)[\tilde{\mathbf{x}}]_i = h(\lambda_i)\mathbf{u}_i^\top \mathbf{x} \, . \tag{9}$$

Therefore, the graph filter modifies the contribution of the i-th component $\mathbf{u}_i$ to the output via the function $h : \mathbb{R} \mapsto \mathbb{R}$ evaluated on the eigenvalue λ_i. Accordingly, $h(\lambda_i)$ is known as the *frequency*

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

response of the graph filter $\mathbf{H}(\cdot)$ at frequency λ_i [45]. In supervised ML settings, the filter taps form the *learnable* parameters that are estimated from data. To reconcile these GSP concepts with the brain age gap prediction task at hand, we note that graph filter taps (and the resulting frequency response) can qualitatively capture the impact of the training procedure, thus, making Principle 2 actionable; see also 'Enter VNNs: GNNs with covariance graphs for structural MRI'. While the capacity of a graph filter is limited to learning *linear* operators, they are the key ingredients in GNNs – the subject dealt with next.

Graph neural networks. GNNs are learnable parametric architectures for *nonlinear* information processing, which are built from graph filter primitives. A *graph perceptron* is constructed by feeding the output of the graph filter through a pointwise nonlinear activation function $\sigma(\cdot)$ (e.g., $\mathsf{ReLU}, \tanh$), that satisfies $\sigma(\mathbf{d}) = [\sigma(d_1), \ldots, \sigma(d_M)]^\top$ for $\mathbf{d} = [d_1, \ldots, d_M]^\top$. Hence, the output of a single layer GNN with input $\mathbf{x}$ is given by $\mathbf{z} = \sigma(\mathbf{H}(\mathbf{A})\mathbf{x})$. For an L-layer GNN, let $\mathbf{H}_\ell(\mathbf{A})$ be the graph filter in layer ℓ and $\mathcal{H}_\ell$ the corresponding set of filter taps. A multilayer (deep) GNN can simply be formed by concatenating individual graph perceptrons, such that the recursive relationship between the input $\mathbf{x}_{\ell-1}$ and the output $\mathbf{x}_\ell$ at the ℓ-th layer is $\mathbf{x}_\ell = \sigma(\mathbf{H}_\ell(\mathbf{A})\mathbf{x}_{\ell-1}),$ for $\ell \in \{1, \ldots, L\}$, where $\mathbf{x}_0$ is the input $\mathbf{x}$.

The expressive power of a GNN architecture can be further enhanced by incorporating multiple input multiple output (MIMO) processing at every layer; see [42] for additional details we omit here to avoid introducing unnecessarily cumbersome notation. In any case, a multilayer GNN architecture capable of MIMO processing is henceforth denoted as $\Psi(\mathbf{x}; \mathbf{A}, \mathcal{H})$, where the set of filter taps $\mathcal{H}$ captures the full span of its architecture. We also write $\Psi(\mathbf{x}; \mathbf{A}, \mathcal{H})$ to denote the output at the GNN's final layer, which is the GNN representation of input $\mathbf{x}$. The output $\Psi(\mathbf{x}; \mathbf{A}, \mathcal{H})$ is typically succeeded by a readout function that maps it to the desired inference outcome.

> The theoretical and operational properties that make GSP appealing to neuroimaging data analysis begin to take shape at the level of the graph filter itself.

Stability of graph filters and GNNs. Interestingly, graph filter outputs are *stable* to various abstract perturbations of $\mathbf{A}$. More formally, $\|\mathbf{H}(\mathbf{A}) - \mathbf{H}(\mathbf{A} + \delta\mathbf{A})\|$ is provably bounded for a controlled $\delta\mathbf{A}$, provided the frequency response $h(\lambda)$ is sufficiently smooth (so-termed Lipschitz conditions) [42]. This property suggests that ML models using graph filters may provide reproducible outcomes, which is relevant to reproducibility in network neuroscience when brain graphs are estimated from different datasets (or sample sizes). Notably, the stability of graph filters readily extends to GNNs; $\|\Psi(\mathbf{x}; \mathbf{A}, \mathcal{H}) - \Psi(\mathbf{x}; \mathbf{A} + \delta\mathbf{A}, \mathcal{H})\|$ is bounded under similar mild Lipschitz conditions on the constituent graph filters.

Transferability of graph filters and GNNs. A given graph filter $\mathbf{H}(\cdot)$, i.e., with fixed filter taps, can be *transferred* to process datasets of arbitrary dimensionalities. Indeed, the same polynomial function $\mathbf{H}(\cdot)$ can be evaluated on matrices of any size. Consider another M'-dimensional graph signal $\mathbf{x}' = [x'_1, x'_2, \ldots, x'_{M'}]^\top$ associated with a graph representation $\mathbf{A}'$ of size $M' \times M'$. In this scenario, the filter taps $\mathbf{h}$ in (5) can be reused to generate another output $\mathbf{z}' = \sum_{k=0}^{K} h_k \mathbf{A}'^k \mathbf{x}' = \mathbf{H}(\mathbf{A}')\mathbf{x}'$, now a vector of length M'. This property readily extends to GNNs based on graph filters, where the same GNN can be used to generate nodal representations from different datasets of distinct dimensionalities. This transferability property supports Principle 4 on 'Generalizability beyond specific dimensionality of data'.

Granted, the success of transference will be measured by the consistency in performance attained across datasets curated according to different brain atlases. Hence, studying the convergence between the graph filter outputs $\mathbf{z}$ and $\mathbf{z}'$ will be central to the theoretical characterization of successful transference.

The convergence between $\mathbf{z}$ and $\mathbf{z}'$ can be formalized by considering the continuous approximations of these discrete objects. Specifically, given an M-dimensional vector $\mathbf{x} = [x_1, \ldots, x_M]^\top$, we can define a continuous representation of $\mathbf{x}$ as a function $y_{\mathbf{x}} : [0,1] \mapsto \mathbb{R}$, such that, $y_{\mathbf{x}}(a) = x_i$ for $a \in \mathcal{U}_i$, where $\mathcal{U}_i$ is a pre-defined subinterval of $[0,1]$ associated with the i-th element of $\mathbf{x}$. Similarly, we can map the matrix $\mathbf{A}$ to a compact set $[0,1]^2$ via $W_{\mathbf{A}} : [0,1]^2 \mapsto \mathbb{R}$, where we have $W_{\mathbf{A}}(a,b) = [\mathbf{A}]_{ij}$ for $a \in \mathcal{U}_i$ and $b \in \mathcal{U}_j$. Note that we can recover $\mathbf{x}$ from $y_{\mathbf{x}}$ and vice-versa (similarly for $\mathbf{A}$ and $W_{\mathbf{A}}$). Hence, for graph signals $\mathbf{x}$ and $\mathbf{x}'$ consisting of M and M' elements, respectively, the closeness of their continuous representations $y_{\mathbf{x}}$ and $y_{\mathbf{x}'}$, i.e., $\|y_{\mathbf{x}} - y_{\mathbf{x}'}\|$ can be used to quantify the similarity between graph signals of different lengths. This observation also extends to the comparison between matrices $\mathbf{A}$ and $\mathbf{A}'$. By leveraging the theory of graphons as limit objects of graphs (i.e., when $M \to \infty$) [46], the convergence between filter outputs $\mathbf{z}$ and $\mathbf{z}'$ via their respective continuous approximations $y_{\mathbf{z}}$ and $y_{\mathbf{z}'}$ was established using so-termed graphon signal processing [42]. Specifically, under smoothness conditions on the graph filter (i.e., the variation between the frequency responses $h(\lambda_i)$ and $h(\lambda_j)$ is bounded as $|h(\lambda_i) - h(\lambda_j)| \leq \vartheta |\lambda_i - \lambda_j|$ for some $\vartheta > 0$ and any pair (λ_i, λ_j)) and the assumption that the continuous approximations $W_{\mathbf{A}}$ and $W_{\mathbf{A}'}$ are part of a converging sequence, the distance $\|y_{\mathbf{z}} - y_{\mathbf{z}'}\|$ vanishes at the rate of $\frac{1}{\sqrt{M}} + \frac{1}{\sqrt{M'}}$ for a graph filter instantiated on graphs $\mathbf{A}$ and $\mathbf{A}'$ of sizes M and M', respectively.

In summary, our GSP-friendly exposition of GNNs revealed that learned representations are intimately tied to the graph eigenvectors, thus, lending some notion of explainability to information processing with GNNs. Moreover, the stability and transferability properties of GNNs support their generalizability to diverse settings. It is precisely in this sense that GSP provides the appropriate substrate to develop the necessary mathematical principles identified in 'Adding mathematical depth to brain age gap prediction'.

Enter VNNs: GNNs with covariance graphs for structural MRI

Data-driven graphs are ubiquitous in neuroimaging data analysis. The anatomical covariance matrix estimated from features derived from structural MRI is a prominent example [47]. Noteworthy contributions on morphometric similarity networks have generalized anatomical covariances to include multiple information modalities within structural MRI [48], with impact to brain age gap prediction [49].

There are critical theoretical gaps in the existing theoretical and empirical properties of GNNs outlined in 'GSP and GNNs: An overview', which are agnostic to the nuanced spatial and spectral characteristics inherent to a data-driven graph. In particular, successful practical adoption of GNNs instantiated on covariance matrices is contingent on a refined mathematical understanding of their properties. In this section, we bridge this gap by surveying the mathematical foundations of data analysis when data-driven covariance graphs are used in GNNs. These architectures are known as coVariance neural networks (VNNs) [40], [50], [51], and we use the terminology 'coVariance filter' to refer to a graph filter implemented on a covariance matrix. Importantly, we discuss the implications of these results on the brain age prediction task deal with here.

Covariance matrix. Covariance matrices are fundamental data structures within multivariate data analysis, that encode statistical dependencies between different pairs of features in a dataset. Our perspective is to view a covariance matrix as a graph representation of a multivariate dataset consisting of n random, independent and identically distributed (i.i.d.) data samples $\mathbf{x}_i \in \mathbb{R}^M, \forall i \in \{1, \ldots, n\}$. In our neuroimaging setting, the data samples are the anatomical features, where M is the number of brain regions of interest (Fig. 5). The empirical covariance matrix is estimated from samples as

$$\hat{\mathbf{C}}_n \triangleq \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\top} , \tag{10}$$

where $\bar{\mathbf{x}}$ is the sample mean across n samples. When samples correspond to cortical brain features, the anatomical covariance matrix is the mathematical construct encoding pairwise statistical interdependencies of brain atrophy across brain regions. Figure 5 illustrates the anatomical features obtained from structural MRI scans (the graph signals) and the process of estimating the anatomical covariance matrix that is used as a graph representation of brain anatomy. From a statistical perspective, the sample covariance matrix is a statistical estimate of the true covariance matrix (also referred to as ensemble covariance matrix) $\mathbf{C}$. This ensemble covariance matrix $\mathbf{C}$ is determined from an $M-$dimensional random vector $\mathbf{x} \in \mathbb{R}^M$ as $\mathbf{C} \triangleq \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{x}])(\mathbf{x} - \mathbb{E}[\mathbf{x}])^{\top}]$. Although the graph filters and GNNs can be analogously defined on both $\mathbf{C}$ and $\hat{\mathbf{C}}_n$, the ensemble covariance matrix $\mathbf{C}$ cannot be observed directly. Thus, in practice we use noisy sample-based statistical estimates and their quality relative to the ensemble counterparts is governed by matrix perturbation theory [52]. This observation also extends to the eigenspectrum of $\hat{\mathbf{C}}_n$ and $\mathbf{C}$. Specifically, consider the eigendecomposition $\mathbf{C} = \mathbf{V}\mathbf{\Phi}\mathbf{V}^{\top}$, where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_M]$ is the $M \times M$ matrix of eigenvectors and $\mathbf{\Phi} = \mathsf{diag}(\phi_1, \ldots, \phi_M)$ is the diagonal matrix of eigenvalues $\phi_1 \geq \phi_2 \geq \cdots \geq \phi_M \geq 0$. Then, the eigendecomposition of the sample covariance matrix is $\hat{\mathbf{C}}_n = \hat{\mathbf{V}}\hat{\mathbf{\Phi}}\hat{\mathbf{V}}^{\top}$, where its matrix of eigenvectors $\hat{\mathbf{V}} = [\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_M]$ are statistical estimates of $\mathbf{V}$ and the eigenvalues $\hat{\mathbf{\Phi}} = \mathsf{diag}(\hat{\phi}_1, \ldots, \hat{\phi}_M)$ are statistical estimates of $\mathbf{\Phi}$. Matrix $\hat{\mathbf{\Phi}}$ is a perturbed version of $\mathbf{\Phi}$.

Case study 1 demonstrated that a reduction in cortical thickness is characteristic of brain atrophy, which manifests in both healthy aging and neurodegeneration. This implies that the correlation structure in the anatomical covariance matrix will be distorted when specific brain regions exhibit accelerated brain atrophy due to neurodegeneration, relative to that of the healthy cohort. Based on this discussion, we contend that a GNN with anatomical features as inputs and the anatomical covariance matrix as the graph provides a suitable framework for the Δ-Age prediction pipeline. These VNN models are discussed next.

CoVariance neural networks and links with PCA. The eigenspectrum of the covariance matrix implicitly captures the structure of a dataset via the *principal components*, and said structure can be exploited via the PCA transform [53]. PCA-regression has been integrated into brain age gap prediction pipelines dating back to the first studies in the field [23]. Interestingly, a coVariance filter draws similarities with the PCA transform [40, Theorem 1]. This connection follows directly from (7) and (8). Specifically, the output of the graph filter instantiated on $\hat{\mathbf{C}}_n$ (i.e., a coVariance filter) depends on the projection of anatomical features onto the covariance eigenvectors – the principal components $\hat{\mathbf{v}}_i^{\top}\mathbf{x}$.

When it comes to qualitative assessment for neuroimaging data analysis, the equivalence between PCA and coVariance filters suggests our argument in 'GSP and GNNs: An overview' can be restated

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

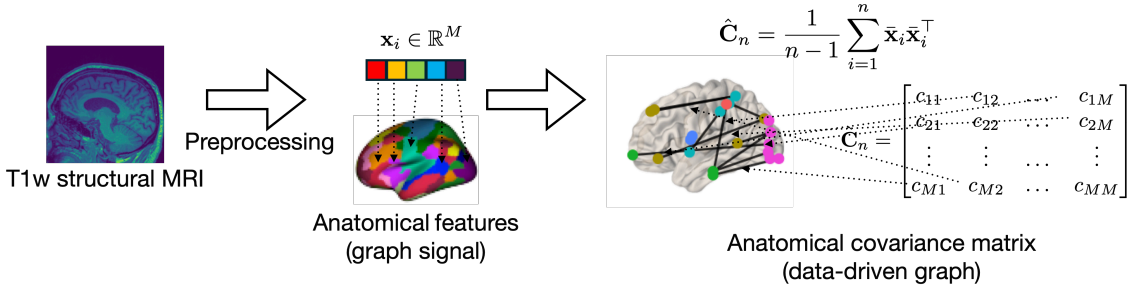

Fig. 5. Extracting graph signals and a data-driven graph from structural MRI. Pre-processing of structural MRI yields anatomical features across the brain cortex. These anatomical features have a vector representation $\mathbf{x}$ of length M, each element of which corresponds to a distinct brain region. From a dataset of anatomical features from n individuals, we can estimate the anatomical covariance matrix $\hat{\mathbf{C}}_n \in \mathbb{R}^{M\times M}$. Here, the shorthand notation $\bar{\mathbf{x}}_i$ stands for $(\mathbf{x}_i - \bar{\mathbf{x}})$ in (10). The anatomical covariance matrix comprises the graph representation of brain anatomy, with its off-diagonal elements characterizing the correlation between anatomical features associated with different brain regions.

as follows: when trained to predict age, a coVariance filter learns specific ways to exploit the principal components of the anatomical covariance matrix. The observations here provide a neat link between the computational module in a deep learning model and a PCA-based feature extractor. Thus, at least in part, VNNs achieve their learning objective (predicting age here) by exploiting the principal components of the anatomical covariance matrix in a judicious, data-driven manner.

To exhibit reproducibility across independent datasets, it is critical that coVariance filters and VNNs be stable to stochastic perturbations in $\hat{\mathbf{C}}_n$ relative to $\mathbf{C}$.

PCA-driven approaches are prone to unstable or irreproducible inference outcomes as a result of stochastic perturbations in the covariance eigenspectrum due to small changes in the dataset (e.g., by addition or removal of a few samples) [54]. VNNs, however, overcome such irreproducibility pitfalls.

Stability of VNNs. The deviation between the outputs of coVariance filters or VNNs for $\hat{\mathbf{C}}_n$ relative to $\mathbf{C}$ is bounded if the frequency response of the coVariance filter is sufficiently smooth in the Lipschitz sense. This stability result, that we informally state next, implies reproducibility of age prediction outcomes when using a VNN model, unlike comparable PCA-driven approaches (Fig. 6).

Theorem 1 (Stability of coVariance Filters and VNNs (Informal) [40]): Consider a random vector $\mathbf{x} \in \mathbb{R}^M$, such that, $\|\mathbf{x}\| \leq 1$, and its corresponding ensemble covariance matrix $\mathbf{C} = \mathbb{E}[(\mathbf{x}-\mathbb{E}[\mathbf{x}])(\mathbf{x}-\mathbb{E}[\mathbf{x}])^\top]$. For a sample covariance matrix $\hat{\mathbf{C}}_n$ formed using n i.i.d. realizations of $\mathbf{x}$, if the frequency response satisfies $|h(\phi_i) - h(\phi_j)| \leq \varsigma |\phi_i - \phi_j|$ for an appropriate $\varsigma > 0$, the following holds with high probability:

$$\left\| \mathbf{H}(\hat{\mathbf{C}}_n) - \mathbf{H}(\mathbf{C}) \right\| \leq \alpha_n, \tag{11}$$

where α_n scales as $\mathcal{O}(1/n^{1/2-\varepsilon})$ for some $\varepsilon \in (0, 1/2)$. Further, for a VNN $\Psi(\cdot;\cdot,\mathcal{H})$ of depth L and F outputs per MIMO layer, if the pointwise non-linearity $\sigma(\cdot)$ satisfies $|\sigma(a) - \sigma(b)| \leq |a-b|$, then

$$\|\Psi(\mathbf{x};\hat{\mathbf{C}}_n,\mathcal{H}) - \Psi(\mathbf{x};\mathbf{C},\mathcal{H})\| \leq LF^{L-1}\alpha_n. \tag{12}$$

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

The eigenvalues $\{\hat{\phi}_1, \ldots, \hat{\phi}_M\}$ of the sample covariance matrix $\hat{\mathbf{C}}_n$ are likely to be perturbed relative to the eigenvalues $\{\phi_1, \ldots, \phi_M\}$ of $\mathbf{C}$. For close eigenvalues of $\mathbf{C}$, the corresponding estimates in $\hat{\mathbf{C}}_n$ may not maintain consistent ordering (in terms of magnitude) with high probability. Hence, a traditional PCA-regression approach is highly vulnerable to irreproducibility when $\hat{\mathbf{C}}_n$ is perturbed. However, this concern is mitigated by VNN-based information processing as the filter response $h(\lambda)$ exhibits limited variability for eigenvectors associated with close eigenvalues (see the assumption in Theorem 1).

Improved stability of coVariance filters and VNNs relative to PCA: An age prediction task

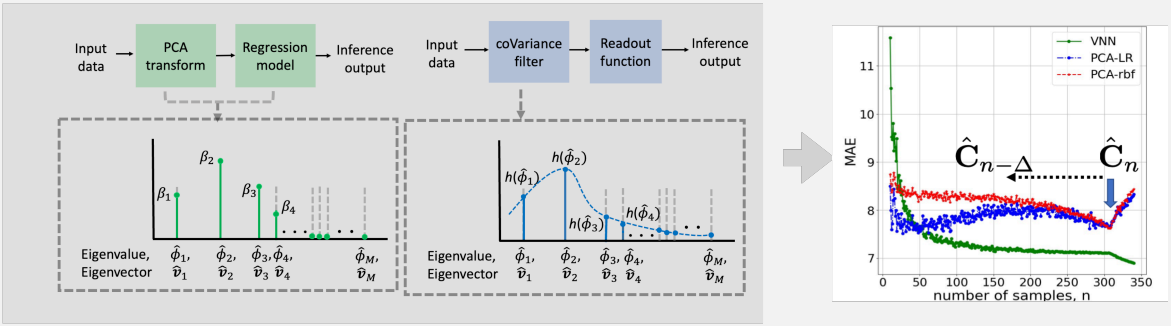

Fig. 6. PCA-regression versus a coVariance filter-driven learning approach to regression. PCA-regression explicitly assigns the importance to specific eigenvectors of the sample covariance matrix $\hat{\mathbf{C}}_n$. The coVariance filter implicitly determines the influence of specific eigenvectors of $\hat{\mathbf{C}}_n$ to the learning task via the frequency response $h(\hat{\phi}_i)$ for eigenvalues $\hat{\phi}_i$ in $\hat{\mathbf{\Phi}}$. As a consequence of Theorem 1, the regression performance of the VNN on the age prediction task using cortical thickness features is consistent even when the sample covariance matrix is perturbed (represented by $\hat{\mathbf{C}}_{n-\Delta}$) relative to the one used for training ($\hat{\mathbf{C}}_n$, marked with a blue arrow) [40]. In contrast, PCA-regression with a linear kernel (PCA-LR) or a radial basis function kernel (PCA-rbf) exhibit instability when the principal components are re-evaluated using perturbed sample covariance matrices, while retaining the same regression weights.

Remark 1 (Sparsifying the anatomical covariance matrix in VNNs): Our discussion has so far implicitly assumed that the number of samples n is large enough for $\hat{\mathbf{C}}_n$ to be a reasonably accurate approximation of $\mathbf{C}$. However, *high-dimensional* neuroimaging settings are characterized by small sample sizes, that will adversely affect estimation of the anatomical covariance matrix. This predicament may create memory inefficiency and computational challenges, especially when the true covariance matrix is sparse but $\hat{\mathbf{C}}_n$ is dense due to spurious correlations. To address such challenges, *sparse* VNNs have been proposed to attain better quality covariance matrices while preserving the stability of VNNs [50]. Sparse VNNs implement principled hard and soft-thresholding strategies to filter out spurious correlations.

Transferability of VNNs. VNNs also inherit the GNN property of transference across datasets of different dimensionalities. This makes VNNs compatible with Principle 4 for Δ-Age prediction. The ensuing discussion introduces the mathematical principles behind 'successful' transference, i.e., when a VNN retains its performance (without retraining) after being deployed to process another dataset of distinct dimensionality. To further exemplify this desirable property and its practical impact, Fig. 11 illustrates the transferability of VNNs in the context of brain age gap prediction.

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

Once more, the theoretical groundwork relies on continuous approximations of discrete objects but now in the VNN setting [55], mimicking the ideas for general GNNs we briefly outlined in 'GSP and GNNs: An Overview'. To ground the abstractions, the M partitions of the interval $[0, 1]$ to generate $y_{\mathbf{x}}$ can be interpreted as a partition of the brain cortex into M regions [55]. The measures of the i-th interval in $y_{\mathbf{x}}$ and the i-th diagonal block in $W_{\mathbf{C}}$ are chosen to be proportional to the marginal variance $[\mathbf{C}]_{ii}$; see also [55] for technical details. The following theorem establishes transference of a VNN with parameters $\mathcal{H}$ between datasets of M_1 and M_2 features. To state the result, let y_{M_1} and y_{M_2} denote the continuous approximations of $\Psi(\mathbf{x}_{M_1}; \mathbf{C}_{M_2}, \mathcal{H})$ and $\Psi(\mathbf{x}_{M_2}; \mathbf{C}_{M_2}, \mathcal{H})$, respectively. The notation $\mathbf{x}_M$ and $\mathbf{C}_M$ explicitly emphasizes that the number of features is M.

Theorem 2 (Transference of VNNs (Informal) [55]): Consider two datasets of M_1 and M_2 features and a VNN $\Psi(\cdot; \cdot, \mathcal{H})$ consisting of L layers and F outputs per MIMO layer. If the continuous approximations $W_{\mathbf{C}_{M_1}}$ and $W_{\mathbf{C}_{M_2}}$ are close and part of a converging sequence to a suitable graphon limit, and the continuous approximations $y_{\mathbf{x}_{M_1}}$ and $y_{\mathbf{x}_{M_2}}$ of the inputs are sufficiently close, then the continuous approximations of the VNN outputs $\Psi(\mathbf{x}_{M_1}; \mathbf{C}_{M_1}, \mathcal{H})$ and $\Psi(\mathbf{x}_{M_2}; \mathbf{C}_{M_2}, \mathcal{H})$ converge in the sense

$$\|y_{M_1} - y_{M_2}\|_2 = \mathcal{O}\left(\frac{1}{M_1^{3\zeta/2-1}} + \frac{1}{M_2^{3\zeta/2-1}}\right), \quad \text{for some constant } \zeta \in (2/3, 1]. \tag{13}$$

Theorem 2 summarizes the foundational principle behind the transference of VNNs, with tangible impacts to neuroimaging data analysis. In fact, the graphon limit can be viewed as the brain cortex, over which finite-dimensional neuroimaging datasets are sampled. The sampling design is determined by the brain atlas used to divide the cortex into regions of interest. The key upshot of Theorem 2 is to justify that VNNs are capable of processing datasets curated according to different brain atlases; see also Fig. 11.

Transferability of VNNs across multiscale datasets for age prediction

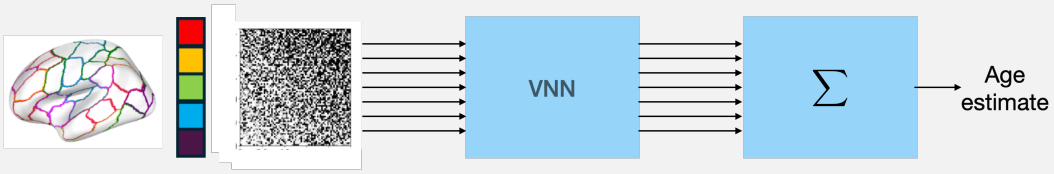

Fig. 7. VNN model schematic designed for age-prediction with transference across datasets of different dimensionalities (Fig. 5). The readout function is the unweighted mean.

TABLE II

TRANSFERABILITY FOR AGE PREDICTION TASK ACROSS MULTISCALE DATASETS (MAE FOR VNN REGRESSION OUTPUTS WITH RESPECT TO THE GROUND TRUTH) [55]

Training \ Testing	100 features	300 features	500 features
100 features	5.39 ± 0.084	5.5 ± 0.101	5.61 ± 0.132
300 features	5.39 ± 0.193	5.41 ± 0.167	5.47 ± 0.169
500 features	5.43 ± 0.2	5.38 ± 0.15	5.4 ± 0.169

The transference of VNN-based regression models across datasets of different dimensionalities is

facilitated by setting the readout function to be the unweighted mean (as illustrated in Fig. 7). For the age prediction task on the same healthy population, here we consider multiscale cortical thickness datasets with $M = 100$, 300, and 500 features spanning the entire brain cortex (curated according to Schaefer's brain atlas) [55]. Table 2 reports the regression performance of three VNNs trained on the datasets with $M = 100$, 300, and 500 features along the diagonal. The performance after transference of the VNNs to a test dataset of different dimensionality is tabulated in the off-diagonal elements. For instance, the element corresponding to the row associated with '100 features' and the column associated with '300 features' indicates the MAE performance for a VNN trained on the dataset with 100 features and transferred to the dataset with 300 features.

TOWARDS EXPLAINABLE BRAIN AGE GAP PREDICTION FROM STRUCTURAL MRI

In this section, we close the loop by blending the various technical results and observations from 'GSP Foundations and Neuroimaging Data Analysis' to construct a workflow for brain age gap prediction that meets the Principles 1-4. We assume that structural MRI scans are pre-processed with standard pipelines to derive anatomical features [4]. The key modules of this workflow can be summarized as follows.

ML model for Δ-Age prediction. A VNN is selected as the regression model and the anatomical covariance matrix $\hat{\mathbf{C}}_n$ is estimated from the anatomical features $\mathbf{x}$ of the healthy population. The covariance matrix $\hat{\mathbf{C}}_n$ remains fixed in the Δ-Age prediction pipeline. The *readout* function of the VNN model is chosen to be an unweighted mean function. Hence, the age estimate $\hat{y}$ is formed by aggregating the learned representations in $\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})$ (Fig. 7) to yield

$$\hat{y} = \frac{1}{M} \sum_{j=1}^{M} [\mathbf{p}_{\mathbf{x}}]_j \ , \quad \text{where} \quad \mathbf{p}_{\mathbf{x}} = \frac{1}{F} \sum_{f=1}^{F} [\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})]_f \tag{14}$$

and F is the width of the VNN at its final layer. The VNN model is pre-trained to predict the chronological age of a healthy cohort. Δ-Age is derived using the steps described in 'How is brain age gap evaluated?'. The estimates $\hat{y}$ across the dataset are further corrected for age bias to yield the brain age estimate $\hat{y}_{\mathsf{B}}$ according to (2), e.g., [15], which further provides the estimate of Δ-Age in (3).

The information gleaned by the VNN model leading to a higher Δ-Age is embedded in the learned representation $\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})$. Traces of accelerated aging in the anatomical features $\mathbf{x}$ of an individual with neurodegeneration are encoded as deviations in $\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})$ relative to what is expected of a healthy individual. These differences propagate downstream through the age estimate $\hat{y}$ to yield a larger Δ-Age. This transparency on the role of the VNN model for Δ-Age prediction would had been lost if we used a learnable readout function (e.g., a multilayer perceptron). In this hypothesized case, the weights of the VNN and the readout function would *collectively* contribute to the age estimate $\hat{y}$, leading to conceptual ambiguities. As we proceed in this section, it will become clear how a VNN with an unweighted mean readout function enables a qualitative assessment of Δ-Age, and seamless transference of the brain age gap prediction pipeline across datasets of different dimensionalities.

Anatomic interpretability of Δ-Age. The anatomic characterization of $\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})$, how this representation leads to the age estimate $\hat{y}$ and subsequently, Δ-Age, are key to establishing the desirable *anatomic*

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

interpretability of Δ-Age. To this end, note that the vector $\mathbf{p_x}$ is obtained by aggregating across the width of $\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})$ in (14), such that, the age estimate $\hat{y}$ is the mean of the elements in $\mathbf{p_x}$. Due to the coVariance filters in the VNN, we can interpret (14) to state that $\mathbf{p_x}$ is formed by transforming the input anatomical features $\mathbf{x}$ according to the covariance matrix $\hat{\mathbf{C}}_n$, where the learnable parameters $\mathcal{H}$ of the VNN encode the information about healthy aging. When projected on the brain atlas, $\mathbf{p_x}$ encodes brain 'regional contributions' to the predicted output $\hat{y}$. For instance, $[\mathbf{p_x}]_j$ is the contribution of region j.

Accelerated aging, as captured by Δ-Age, could be hypothesized to be an aggregated effect of anomalous contributions from certain biologically plausible brain regions. Hence, if the representation $\Psi(\mathbf{x}; \hat{\mathbf{C}}_n, \mathcal{H})$ encodes the information about accelerated aging, we anticipate that certain elements of $\mathbf{p_x}$ will exhibit a 'larger contribution' to the age estimate $\hat{y}$. The *anatomical signatures* of Δ-Age for a neurodegenerative condition can be revealed via group comparisons of appropriately defined statistics of $\mathbf{p_x}$, between the cohort with the neurodegenerative condition and a healthy cohort. Direct comparisons between different populations are prone to bias due to differences in the respective chronological age distributions. To mitigate this bias and better capture accelerated aging in the disease cohort during group-level analyses, we define the 'regional residual' statistic for anatomical feature j (or brain region represented by feature j in this case) with respect to the VNN output $\hat{y}$ at the regional level as

$$[\mathbf{r}]_j \triangleq [\mathbf{p_x}]_j - \hat{y} \; . \tag{15}$$

To understand the contribution of elevated regional residuals to higher Δ-Age for a cohort with accelerated aging, consider a toy example with two individuals of the same chronological age y. Suppose that one belongs to the disease group, the other to the healthy cohort. From (2), their corresponding VNN outputs (denoted by $\hat{y}_{\mathsf{D}}$ for the individual in the disease cohort and $\hat{y}_{\mathsf{HC}}$ for the individual in the healthy cohort) are corrected for age-bias using the same term $\omega y + \varrho$. Hence, Δ-Age for the individual in the disease cohort will be highest only if the VNN prediction $\hat{y}_{\mathsf{D}}$ exceeds $\hat{y}_{\mathsf{HC}}$. Since the VNN predictions $\hat{y}_{\mathsf{D}}$ and $\hat{y}_{\mathsf{HC}}$ are proportional to the unweighted aggregations of the regional level estimates [see (14)], $\hat{y}_{\mathsf{D}} > \hat{y}_{\mathsf{HC}}$ can be a direct consequence of a subset of regional residuals [see (15)] being robustly elevated in the disease group relative to the healthy cohort. If the individuals have different chronological ages, the age-bias correction will remove any age-related confounding in the differences in distributions of Δ-Age.

Based on the arguments above, the brain regions contributing to higher Δ-Age in neurodegeneration can be traced to significantly elevated regional residuals in the disease cohort. Anatomical interpretability thus follows via evaluation of group level differences between the regional residuals of the disease and healthy cohorts (with standard tests such as ANCOVA using variables like age and gender as covariates). The elements of the residual vector $\mathbf{r}$ have been shown to exhibit distinct anatomic signatures for Δ-Age under different neurodegenerative conditions [25], [26], which could be used for 'fingerprinting'.

Explaining regional residuals. Our discussion in 'GSP Foundations for Neuroimaging Data Analysis' revealed that learning with VNNs can be understood, at least in part, as a process that implicity exploits the eigenvectors of $\hat{\mathbf{C}}_n$. Since the regional residuals are derived directly from the VNN representations, the anomalous behavior in neurodegeneration could be captured (and explained) in terms of how the VNN selectively leveraged said eigenvectors across the different population groups. To this end, for an

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

individual with regional residual vector $\mathbf{r}$, we consider the inner product $\bar{\mathbf{r}}^\top \hat{\mathbf{v}}_i$ with eigenvector $\hat{\mathbf{v}}_i$ as metric, where $\bar{\mathbf{r}}$ is the normalized version of $\mathbf{r}$, such that, $\|\bar{\mathbf{r}}\|_2 = 1$.

Notably, the inner product metric $\bar{\mathbf{r}}^\top \hat{\mathbf{v}}_i$ closely resembles the GFT (here coVariance Fourier transform) in (9) if the ML model were a coVariance filter. In this case and from (9), it follows that $\bar{\mathbf{r}}^\top \hat{\mathbf{v}}_i$ is a composite metric that combines the frequency response of the coVariance filter $h(\lambda_i)$ and the alignment between the considered eigenvector and the input data, i.e., $\mathbf{v}_i^\top \mathbf{x}$. The analytical extension of this observation to VNNs is non-trivial. However, since the coVariance filter forms the fundamental computational module in a VNN, we anticipate that the metric $\bar{\mathbf{r}}^\top \hat{\mathbf{v}}_i$ will have different group-level distributions for individuals with neurodegeneration and healthy controls (at least for some eigenvectors). This observation motivates our approach to explainability of Δ-Age in neurodegeneration. The terms interpretability and explainability are sometimes used interchangeably in the field of explainable AI. Here, we refer to 'interpretability' when alluding to anatomical interpretability of Δ-Age. The term 'explainablity' is used in the context of understanding how the statistics that yield anatomical interpretability were derived by the VNN. Establishing explainability in this spirit helps disentangle Δ-Age regardless of whether it aligns with the biological hypotheses. Hence, explainability of Δ-Age provides a promising approach not only to support Δ-Age prediction, but also to diagnose unexpected outcomes from a VNN-driven pipeline (for instance, low Δ-Age in a specific set of individuals affected by neurodegeneration).

Case Study 3: Anatomically interpretable and explainable Δ-Age in AD

In this case study, we demonstrate the integration of VNN-driven modules within the Δ-Age prediction workflow for AD. Figure 8 summarizes the workflow for anatomically interpretable Δ-Age prediction using VNNs.

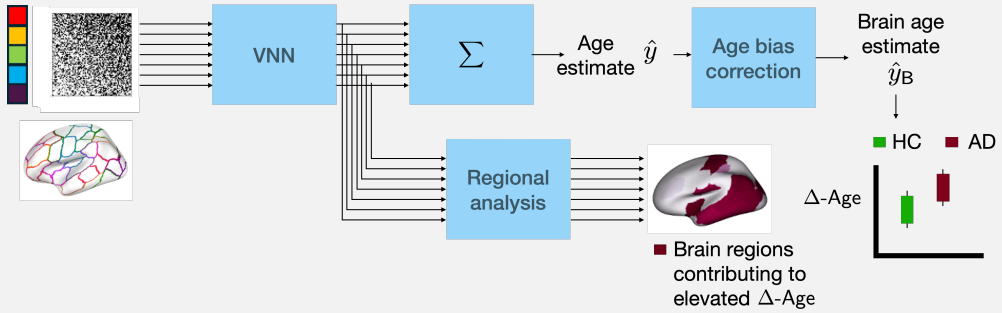

Fig. 8. Workflow for anatomically interpretable Δ-Age prediction using VNNs in the HC and AD cohorts.

VNN model. Recall the ML model in Case Study 2, namely a VNN trained to predict the chronological age of the HC group from the OASIS-3 dataset [17]. The model has $L = 2$ layers, with the first layer consisting of 2 filter taps and the second layer consisting of 6 filter taps. The width is $F = 61$. Overall, the VNN model has $22,570$ learnable parameters. The architecture was determined via a hyperparameter optimization procedure using the Optuna package [56].

Anatomically interpretable Δ-Age. Figure 9a re-illustrates the distributions of Δ-Age for the HC and AD cohorts (see Case Study 2 for additional information). Δ-Age was derived using the same workflow as described in 'How is brain age gap evaluated'? Anatomic interpretability via

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

analyses of regional residuals as described in 'Towards Explainable Brain Age Gap Prediction from Structural MRI' yielded the anatomical map in Fig. 9b. The F-values derived from ANCOVA (age as covariate) from statistically significant group differences (p-value after Bonferroni correction for multiple comparisons < 0.05) in the regional residuals between the HC and AD groups have been projected on the brain surface. Brain regions colored with darker contrast represent the most significant contributors to elevated Δ-Age in the AD group relative to the HC group in Fig. 9a.

Comparing anatomic interpretability with brain atrophy in AD. Case Study 1 revealed that the AD group exhibited larger brain atrophy relative to the HC group (after controlling for age). Moreover, the atrophy was spread across most brain regions in the AD group, being most statistically significant in the bilateral brain regions spanning the temporal lobe, temporo-parietal junction, and entorhinal regions. Notably, the anatomic interpretability of Δ-Age in Fig. 9b aligned with the brain regions exhibiting the most atrophy in Fig. 1d. Since brain atrophy patterns reflect accelerated aging in structural MRI, we expected an alignment between brain atrophy and anatomic interpretability of Δ-Age. Hence, our experiments attested to the fact that Δ-Age in the AD group was indeed driven by atrophy patterns in structural MRI. This finding is challenging to establish reliably using 'black-box' deep learning models. We also note that the anatomic interpretablity supporting elevated Δ-Age in AD (Fig. 9b) was not identical to the brain atrophy patterns in Fig. 1d. Thus, the VNN model refined the information within structural MRI to reveal the key contributors to Δ-Age in AD.

Explaining anatomic interpretability of Δ-Age. In Figure 10a, the bars represent the means of the inner product metrics $\bar{\mathbf{r}}^\top \hat{\mathbf{v}}_i$ calculated between the first 10 eigenvectors of the anatomical covariance matrix and the normalized regional residuals for the AD group. The whiskers in the bar plot in Fig. 10a illustrate the standard deviations of the inner product metrics across the AD group. The results herein revealed that the eigenvectors of the anatomical covariance matrix exhibited non-uniform importance to Δ-Age, with the first four eigenvectors exhibiting the largest relevance to Δ-Age in AD. Furthermore, the group level comparisons of the inner product metrics between AD and HC groups via ANCOVA (with age as covariate) revealed statistically significant differences observed in the inner product metrics for HC and AD groups for various eigenvectors (results for first, second, and sixth eigenvectors are included in Fig. 10b-d. Altogether, the results in Fig. 10 corroborate the overall relevance of eigenvectors to Δ-Age as well as the relative importance of various eigenvectors for Δ-Age in the AD and HC groups.

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

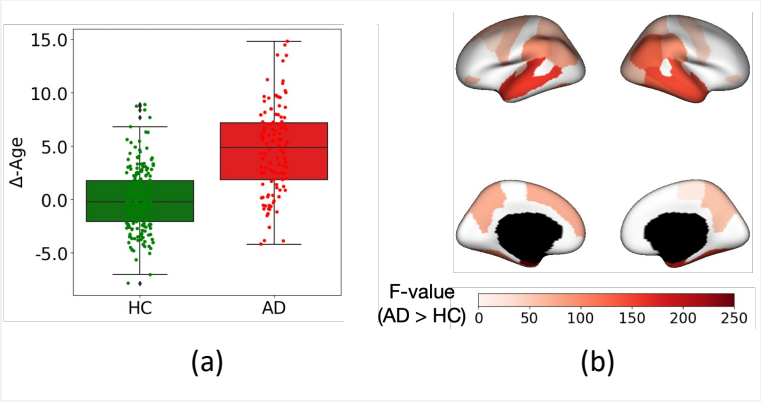

Fig. 9. (a) Δ-Age distributions in the HC and AD groups. (b) Higher Δ-Age in AD relative to HC is anatomically interpreted by the group-level analysis of regional residuals for the HC and AD populations.

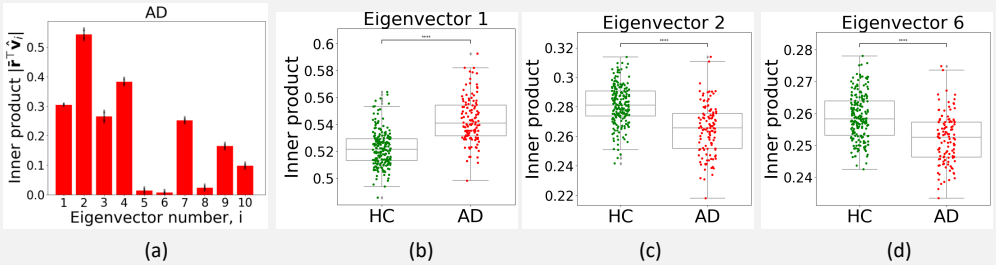

Fig. 10. (a) Mean and standard deviation for the inner product metrics $\bar{\mathbf{r}}^\top \mathbf{v}_i$ for $i \in \{1, \ldots, 10\}$ across the AD group. (b)-(d) The inner product metrics $\bar{\mathbf{r}}^\top \hat{\mathbf{v}}_i$ for eigenvectors 1, 2, and 6 exhibited significant group differences (ANOVA, p-value < 0.05) between the HC and AD cohorts.

Case Study 3 demonstrates how the Δ-Age derived using a VNN model achieves the Principles 1-3 identified earlier. Specifically, by setting the readout function to be an unweighted mean, Δ-Age can be synthesized in terms of VNN-output regional residuals defined at the anatomic level (achieving Principle 1). Elevated Δ-Age in AD vs HC can be traced to regional residual differences at specific brain regions, which are characteristic of AD pathology (such as the medial temporal lobe, among others). The inner product metrics between regional residuals and the eigenvectors of the anatomical covariance matrix revealed how VNNs processed the data for AD, and how this differed from the HC group (thus, achieving Principle 3). Specifically, during pre-training on the healthy population, the VNN model learned to exploit the eigenvectors of the anatomical covariance matrix in a certain way. This led to specific patterns in regional residuals for the HC population (hence, addressing Principle 2). The regional residuals exhibited distinct behavior for specific brain regions in the AD group.

Case Study 4: Transferability of VNNs validates Δ-Age on multiresolution datasets.

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

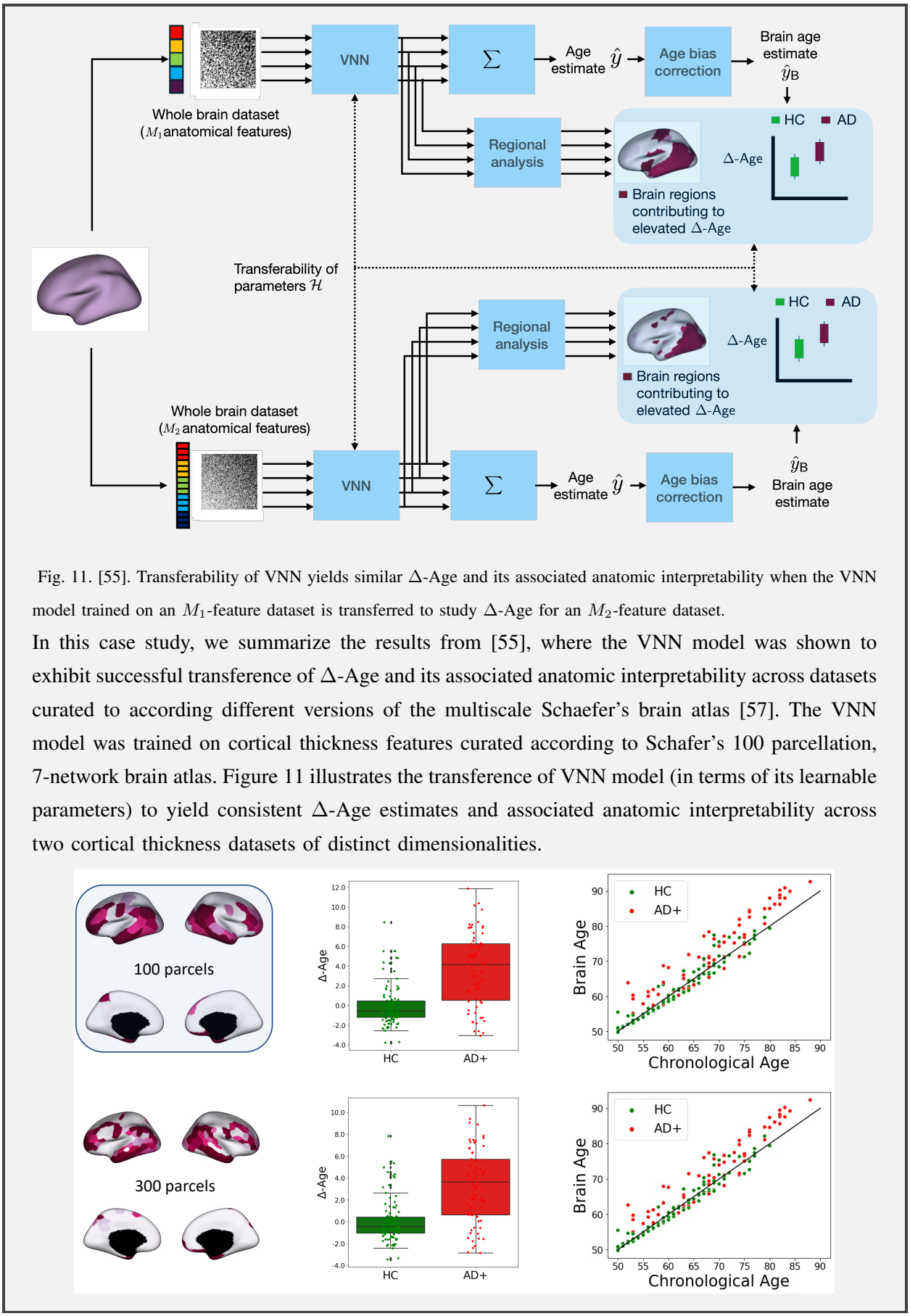

Fig. 11. [55]. Transferability of VNN yields similar Δ-Age and its associated anatomic interpretability when the VNN model trained on an M_1-feature dataset is transferred to study Δ-Age for an M_2-feature dataset.

In this case study, we summarize the results from [55], where the VNN model was shown to exhibit successful transference of Δ-Age and its associated anatomic interpretability across datasets curated to according different versions of the multiscale Schaefer's brain atlas [57]. The VNN model was trained on cortical thickness features curated according to Schafer's 100 parcellation, 7-network brain atlas. Figure 11 illustrates the transference of VNN model (in terms of its learnable parameters) to yield consistent Δ-Age estimates and associated anatomic interpretability across two cortical thickness datasets of distinct dimensionalities.

This article has been accepted for publication in IEEE Signal Processing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MSP.2025.3596731

Fig. 12. [55]. Empirical validation of transference of VNNs across cortical thickness datasets curated according to Schaefer's brain atlas for Δ-Age prediction and associated anatomical interpretability.

Successful transference of VNNs in this context is supported by the theoretical properties of VNNs (Theorem 2 and associated conditions). The anatomical covariance matrices for the datasets curated according to 100-features or 300-features brain atlases can be shown to be part of a converging sequence [55], corroborating the findings in Fig. 12.

Case Study 4 offers empirical support to the theoretical result in Theorem 2. Under certain regularity conditions, the Δ-Age and its associated anatomic interpretability inferred by a specific VNN model can be transferred across neuroimaging datasets of different dimensionalities (thus, achieving Principle 4).

Explainable Δ-Age prediction beyond AD. We have demonstrated various facets of a VNN-driven Δ-Age prediction pipeline by focusing on AD. However, the ML model is broadly applicable as it does not depend on a specific neurodegenerative disease. In fact, the findings in [25] indicate that VNNs can generate an anatomically interpretable Δ-Age in FTD, CBS, and PSP conditions. These neurodegenerative conditions exhibit brain atrophy patterns different from AD and therefore, their anatomical signatures differed from Fig. 9b. For instance, the VNN-driven Δ-Age prediction in FTD revealed both frontal and temporal regions as contributors to FTD in [25], unlike for AD in Fig. 9b, where temporal regions were prominently implicated.

Furthermore, the VNN-driven Δ-Age pipeline can also yield clinical insights into cognitively healthy individuals. Specifically, in [16], Δ-Age was reported to be correlated with plasma neurofilament light chain (NfL) for a cohort of amyloid-positive and cognitively healthy individuals in the ADNI dataset. Amyloid positivity has been linked with accelerated cognitive decline [58] and plasma NfL is a promising blood biomarker of axonal degeneration [59]. Notably, the correlation between Δ-Age and plasma NfL in this cohort was driven by the regional residuals in bilateral entorhinal regions, which are implicated in the early stages of AD pathology [60]. Thus, the discussion here provides promising evidence to support the broader impacts of VNN-driven Δ-Age towards understanding neurodegenerative conditions in various stages of the disease.

Conclusions and Future Outlook

Brain age gap is a promising ML-driven biomarker derived from neuroimaging data, which has not yet been widely adopted in practice due to several methodological obscurities. In this tutorial, we highlighted the major challenges facing prevalent approaches and concluded that performance-driven methods are inadequate for the practical viability of this application. Our focus has been primarily on structural MRI, because it is the most widely adopted neuroimaging modality in clinical applications. Brain age prediction algorithms designed for other neuroimaging modalities exhibit similar shortcomings. In this context, we identified four key mathematical principles that could embellish the practical viability of brain age gap prediction. Broadly, we argued for a shift in focus towards brain age gap instead of brain age, for qualitative (and not performance-driven) assessments of regression models trained on a healthy population, and for the generalizability of Δ-Age to different collections of anatomical features derived from structural MRI.

Hence, an amalgamation of mathematical principles and operational requirements of neuroimaging data analysis is critically needed to address the current limitations facing Δ-Age prediction. To this end, we identified GSP as the key analytical tool to enable principled prediction of Δ-Age. GSP-driven learning architectures benefit from improved interpretability of learning outcomes in terms of spectral representations of the graph structure, as well as much-needed theoretical guarantees on robustness and generalizability. We surveyed a mix of theoretical results on VNNs and case studies to highlight the steps for the principled construction of Δ-Age, its anatomic interpretability, explainability, and generalizability. Admittedly, the robustness of age prediction models to factors such as distribution shifts is key to achieve reproducible outcomes for different neurodegenerative cohorts. A holistic qualitative performance analysis along with a theoretical understanding of the ML model responsible for Δ-Age prediction is much needed.

Looking ahead, brain age gap prediction is a promising tool with a potentially transformative translational impact on digital health and precision medicine. One of the most attractive characteristics of this general approach is its wide applicability to a variety of neurodegenerative conditions. We contend that the underlying ML models possess a characteristic similar to a foundation model, where they can transfer the information learned from healthy aging to yield meaningful biomarkers for various neurodegenerative conditions. Hence, brain age gap prediction algorithms can inform the development of domain-specific foundation models for brain health assessments in the clinic. Our perspective in this article is to also present GSP as a valuable analytical tool in this relatively unexplored context. Extending the principles surveyed in this paper to richer anatomical networks (such as morphometric similarity networks [48]) for brain age gaps prediction is a promising future direction.

Characterizing heterogeneity within disease populations has become increasingly relevant recently, as it can enable targeted interventions and therapies. We argued that GSP-informed architectures facilitate seamless integration of information within anatomical features derived from structural MRI, aging, and the anatomical covariance matrix, to yield representations predictive of accelerated aging in neurodegeneration. More broadly, we believe that GSP principles could be fruitfully leveraged to unveil heterogeneous impacts of neurodegeneration via subtyping of clinically-relevant populations.

In summary, GSP tools hold great promise in becoming one of the dominant analytic paradigms driving ongoing pursuits in digital health and precision medicine, with tangible impacts in the near future. The theoretical foundation of GSP-driven ML models lends unparalleled depth and reliability to their outcomes. This article provides a compelling example of how ML theory can inform major conceptual advancements in the design of data-scientific and application-relevant neuroimaging solutions.

ACKNOWLEDGEMENT

Data used in Case Studies 1 and 2 were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this article. Data collection and sharing for the Alzheimer's Disease Neuroimaging Initiative (ADNI) is funded by the National Institute on Aging (National Institutes of Health Grant U19AG024904). The grantee organization is the Northern California Institute for Research and Education. In the past, ADNI

has also received funding from the National Institute of Biomedical Imaging and Bioengineering, the Canadian Institutes of Health Research, and private sector contributions through the Foundation for the National Institutes of Health (FNIH).

The brain plots were created using the 'fsbrain' package in R [61].

BIOGRAPHIES

Saurabh Sihag is an Assistant Professor in the Department of Electrical and Computer Engineering at the University at Albany. He received his PhD degree in Electrical Engineering from Rensselaer Polytechnic Institute, in 2020. He has previously been the recipient of J. Baliga fellowship and Charles M. Close '62 Doctoral Prize for his doctoral dissertation. His research interests include statistical signal processing, network neuroscience, machine learning, and information theory.

Gonzalo Mateos received his B.Sc. degree in Electrical Engineering from Universidad de la Republica, Montevideo, Uruguay in 2005 and the M.Sc. and Ph.D. degrees in Electrical Engineering from the University of Minnesota, Minneapolis, in 2009 and 2012. Currently, he is a Professor with the Department of Electrical and Computer Engineering, University of Rochester, as well as the Associate Director for Research at the Goergen Institute for Data Science and Artificial Intelligence. He also was an Asaro Biggar Family Fellow in Data Science (2020-23). His research interests lie in the areas of statistical learning from complex data, network science, decentralized optimization, and graph signal processing.

Alejandro Ribeiro received the B.Sc. degree in Electrical Engineering from the Universidad de la República Oriental del Uruguay, Montevideo, Uruguay, in 1998, the M.Sc. and Ph.D. degrees in Electrical Engineering from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, in 2005 and 2007, respectively. Since 2008, he has been with the University of Pennsylvania (Penn), Philadelphia, PA, USA, where he is currently a Professor of Electrical and Systems Engineering. His research interests include the applications of statistical signal processing to the study of networks and networked phenomena, structured representations of networked data structures, graph signal processing, network optimization, robot teams, and networked control.

REFERENCES

[1] P. Arrondo, Ó. Elía-Zudaire, G. Martí-Andrés, M. A. Fernández-Seara, and M. Riverol, "Grey matter changes on brain MRI in subjective cognitive decline: a systematic review," *Alzheimer's Research & Therapy*, vol. 14, no. 1, p. 98, 2022.

[2] S. Przedborski, M. Vila, V. Jackson-Lewis, *et al.*, "Series introduction: Neurodegeneration: What is it and where are we?," *J. Clin. Investig.*, vol. 111, no. 1, pp. 3–10, 2003.

[3] B. T. Wyman, D. J. Harvey, K. Crawford, M. A. Bernstein, O. Carmichael, P. E. Cole, P. K. Crane, C. DeCarli, N. C. Fox, J. L. Gunter, *et al.*, "Standardization of analysis sets for reporting results from ADNI MRI data," *Alzheimer's & Dementia*, vol. 9, no. 3, pp. 332–337, 2013.

[4] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[5] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, vol. 31, no. 3, pp. 968–980, 2006.

[6] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010.

[7] J. Koikkalainen, H. Rhodius-Meester, A. Tolonen, F. Barkhof, B. Tijms, A. W. Lemstra, T. Tong, R. Guerrero, A. Schuh, C. Ledig, *et al.*, "Differential diagnosis of neurodegenerative diseases using structural MRI data," *NeuroImage: Clinical*, vol. 11, pp. 435–449, 2016.

[8] C. López-Otín *et al.*, "The hallmarks of aging," *Cell*, vol. 153, no. 6, pp. 1194–1217, 2013.

[9] R. Peters, "Ageing and the brain," *Postgraduate Medical Journal*, vol. 82, no. 964, pp. 84–88, 2006.

[10] L. Ferrucci, M. Gonzalez-Freire, E. Fabbri, E. Simonsick, T. Tanaka, Z. Moore, S. Salimi, F. Sierra, and R. de Cabo, "Measuring biological aging in humans: A quest," *Aging Cell*, vol. 19, no. 2, p. e13080, 2020.

[11] L. Baecker, J. Dafflon, P. F. Da Costa, R. Garcia-Dias, S. Vieira, C. Scarpazza, V. D. Calhoun, J. R. Sato, A. Mechelli, and W. H. Pinaya, "Brain age prediction: A comparison between machine learning models using region-and voxel-based morphometric data," *Human Brain Mapp.*, vol. 42, no. 8, pp. 2332–2346, 2021.

[12] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector, and G. Montana, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, pp. 115–124, 2017.

[13] F. Farokhian, I. Beheshti, D. Sone, and H. Matsuda, "Comparing CAT12 and VBM8 for detecting brain morphological abnormalities in temporal lobe epilepsy," *Frontiers in neurology*, vol. 8, p. 428, 2017.

[14] J. H. Cole and K. Franke, "Predicting age using neuroimaging: Innovative brain ageing biomarkers," *Trends in Neurosci.*, vol. 40, no. 12, pp. 681–690, 2017.

[15] A.-M. G. de Lange and J. H. Cole, "Commentary: Correction procedures in brain-age prediction," *NeuroImage Clin.*, vol. 26, 2020.

[16] S. Sihag and A. Ribeiro, "Brain age correlates with plasma nfl in amyloid positive individuals," *Alzheimer's & Dementia*, vol. 20, p. e089232, 2024.

[17] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko, *et al.*, "OASIS-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *MedRxiv*, 2019.

[18] S. E. O'Bryant, S. C. Waring, C. M. Cullum, J. Hall, L. Lacritz, P. J. Massman, P. J. Lupo, J. S. Reisch, R. Doody, T. Alzheimer's Research Consortium, *et al.*, "Staging dementia using clinical dementia rating scale sum of boxes scores: a texas Alzheimer's research consortium study," *Archives of Neurology*, vol. 65, no. 8, pp. 1091–1095, 2008.

[19] K. Franke, G. Ziegler, S. Klöppel, C. Gaser, A. D. N. Initiative, *et al.*, "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *Neuroimage*, vol. 50, no. 3, pp. 883–892, 2010.

[20] L. C. Löwe, C. Gaser, and K. Franke, "The effect of the APOE genotype on individual BrainAGE in normal aging, mild cognitive impairment, and Alzheimer's disease," *PloS one*, vol. 11, no. 7, p. e0157514, 2016.

[21] C. Yin, P. Imms, M. Cheng, A. Amgalan, N. F. Chowdhury, R. J. Massett, N. N. Chaudhari, X. Chen, P. M. Thompson, P. Bogdan, *et al.*, "Anatomically interpretable deep learning of brain age captures domain-specific cognitive impairment," *Proc. the National Academy of Sciences*, vol. 120, no. 2, p. e2214634120, 2023.

[22] H. G. Schnack, N. E. Van Haren, M. Nieuwenhuis, H. E. Hulshoff Pol, W. Cahn, and R. S. Kahn, "Accelerated brain aging in schizophrenia: a longitudinal pattern recognition study," *American Journal of Psychiatry*, vol. 173, no. 6, pp. 607–616, 2016.

[23] J. H. Cole, R. Leech, D. J. Sharp, and A. D. N. Initiative, "Prediction of brain age suggests accelerated atrophy after traumatic brain injury," *Annals of Neurology*, vol. 77, no. 4, pp. 571–581, 2015.

[24] C. R. Eickhoff, F. Hoffstaedter, J. Caspers, K. Reetz, C. Mathys, I. Dogan, K. Amunts, A. Schnitzler, and S. B. Eickhoff, "Advanced brain ageing in parkinson's disease is related to disease duration and individual impairment," *Brain Communications*, vol. 3, no. 3, p. fcab191, 2021.

[25] S. Sihag, G. Mateos, and A. Ribeiro, "Explainable brain age gap prediction in neurodegenerative conditions using covariance neural networks," in *Proc. IEEE Intl. Symp. Biomedical Imaging*, 2025.

[26] S. Sihag, G. Mateos, C. McMillan, and A. Ribeiro, "Explainable brain age prediction using covariance neural networks," in *Proc. Conf. Adv. in Neural Inf. Process. Syst.*, 2023.

[27] N. C. Ho, R. A. Bethlehem, J. Seidlitz, N. Nogovitsyn, P. Metzak, P. L. Ballester, S. Hassel, S. Rotzinger, J. Poppenk, R. W. Lam, *et al.*, "Atypical brain aging and its association with working memory performance in major depressive disorder," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 9, no. 8, pp. 786–799, 2024.

[28] R. A. Bethlehem, J. Seidlitz, S. R. White, J. W. Vogel, K. M. Anderson, C. Adamson, S. Adler, G. S. Alexopoulos, E. Anagnostou, A. Areces-Gonzalez, *et al.*, "Brain charts for the human lifespan," *Nature*, vol. 604, no. 7906, pp. 525–533, 2022.

[29] K. Franke and C. Gaser, "Ten years of BrainAGE as a neuroimaging biomarker of brain aging: What insights have we gained?," *Front. Neurol.*, p. 789, 2019.

[30] J. H. Cole, R. E. Marioni, S. E. Harris, and I. J. Deary, "Brain age and other bodily 'ages': implications for neuropsychiatry," *Molecular psychiatry*, vol. 24, no. 2, pp. 266–281, 2019.

[31] J. Lee, B. J. Burkett, H.-K. Min, M. L. Senjem, E. S. Lundt, H. Botha, J. Graff-Radford, L. R. Barnard, J. L. Gunter, C. G. Schwarz, *et al.*, "Deep learning-based brain age prediction in normal aging and dementia," *Nature Aging*, vol. 2, no. 5, pp. 412–424, 2022.

[32] C. Gaser, P. Kalc, and J. H. Cole, "A perspective on brain-age estimation and its clinical promise," *Nature Computational Science*, pp. 1–8, 2024.

[33] M. Azzam, Z. Xu, R. Liu, L. Li, K. Meng Soh, K. B. Challagundla, S. Wan, and J. Wang, "A review of artificial intelligence-based brain age estimation and its applications for related diseases," *Briefings in Functional Genomics*, p. elae042, 2024.

[34] V. M. Bashyam, G. Erus, J. Doshi, M. Habes, I. M. Nasrallah, M. Truelove-Hill, D. Srinivasan, L. Mamourian, R. Pomponio, Y. Fan, *et al.*, "MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14468 individuals worldwide," *Brain*, vol. 143, no. 7, pp. 2312–2324, 2020.

[35] R. J. Jirsaraie, A. J. Gorelik, M. M. Gatavins, D. A. Engemann, R. Bogdan, D. M. Barch, and A. Sotiras, "A systematic review of multimodal brain age studies: Uncovering a divergence between model accuracy and utility," *Patterns*, vol. 4, no. 4, 2023.

[36] A.-K. Dombrowski, M. Alber, C. Anders, M. Ackermann, K.-R. Müller, and P. Kessel, "Explanations can be manipulated and geometry is to blame," in *Proc. Conf. Adv. in Neural Inf. Process. Syst.*, vol. 32, 2019.

[37] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," in *Proc. Conf. Adv. in Neural Inf. Process. Syst.*, vol. 31, 2018.

[38] E. Black, M. Raghavan, and S. Barocas, "Model multiplicity: Opportunities, concerns, and solutions," in *Proc. the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 850–863, 2022.

[39] T. A. Woolsey, J. Hanaway, and M. H. Gado, *The brain atlas: A visual guide to the human central nervous system*. John Wiley & Sons, 2017.

[40] S. Sihag, G. Mateos, C. McMillan, and A. Ribeiro, "coVariance neural networks," in *Proc. Conf. Adv. in Neural Inf. Process. Syst.*, Nov. 2022.

[41] G. Leus, A. G. Marques, J. M. Moura, A. Ortega, and D. I. Shuman, "Graph signal processing: History, development, impact, and outlook," *IEEE Signal Process. Mag.*, vol. 40, no. 4, pp. 49–60, 2023.

[42] L. Ruiz, F. Gama, and A. Ribeiro, "Graph neural networks: architectures, stability, and transferability," *Proc. IEEE*, vol. 109, no. 5, pp. 660–682, 2021.

[43] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neuroscience*, vol. 20, no. 3, pp. 353–364, 2017.

[44] W. Huang, T. A. Bolton, J. D. Medaglia, D. S. Bassett, A. Ribeiro, and D. Van De Ville, "A graph signal processing perspective on functional brain imaging," *Proc. IEEE*, vol. 106, no. 5, pp. 868–885, 2018.

[45] E. Isufi, F. Gama, D. I. Shuman, and S. Segarra, "Graph filters for signal processing and machine learning on graphs," *IEEE Transactions on Signal Processing*, vol. 72, pp. 4745–4781, 2024.

[46] L. Lovász, *Large Networks and Graph Limits*, vol. 60. American Mathematical Soc., 2012.

[47] A. C. Evans, "Networks of anatomical covariance," *Neuroimage*, vol. 80, pp. 489–504, 2013.

[48] J. Seidlitz, F. Váša, M. Shinn, R. Romero-Garcia, K. J. Whitaker, P. E. Vértes, K. Wagstyl, P. K. Reardon, L. Clasen, S. Liu, *et al.*, "Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation," *Neuron*, vol. 97, no. 1, pp. 231–247, 2018.

[49] P. Galdi, M. Blesa, D. Q. Stoye, G. Sullivan, G. J. Lamb, A. J. Quigley, M. J. Thrippleton, M. E. Bastin, and J. P. Boardman, "Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth," *NeuroImage: Clinical*, vol. 25, p. 102195, 2020.

[50] A. Cavallo, Z. Gao, and E. Isufi, "Sparse covariance neural networks," *arXiv:2410.01669*, vol. cs.LG, 2024.

[51] A. Cavallo, M. Sabbaqi, and E. Isufi, "Spatiotemporal covariance neural networks," in *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*, pp. 18–34, Springer, 2024.

[52] A. Loukas, "How close are the eigenvectors of the sample and actual covariance matrices?," in *Proc. Int. Conf. Mach. Learn.*, pp. 2228–2237, 2017.

[53] J. Shlens, "A tutorial on principal component analysis," *arXiv preprint arXiv:1404.1100*, 2014.

[54] I. T. Joliffe and B. Morgan, "Principal component analysis and exploratory factor analysis," *Stat. Methods Med. Res.*, vol. 1, no. 1, pp. 69–95, 1992.

[55] S. Sihag, G. Mateos, C. McMillan, and A. Ribeiro, "Transferability of covariance neural networks," *IEEE J. Sel. Topics Signal Process.*, pp. 1–16, 2024.

[56] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 2623–2631, 2019.

[57] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X.-N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. Yeo, "Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI," *Cerebral Cortex*, vol. 28, no. 9, pp. 3095–3114, 2018.

[58] O. Janssen, W. J. Jansen, S. J. Vos, M. Boada, L. Parnetti, T. Gabryelewicz, T. Fladby, J. L. Molinuevo, S. Villeneuve, J. Hort, *et al.*, "Characteristics of subjective cognitive decline associated with amyloid positivity," *Alzheimer's & Dementia*, vol. 18, no. 10, pp. 1832–1845, 2022.

[59] M. M. Mielke, J. A. Syrjanen, K. Blennow, H. Zetterberg, P. Vemuri, I. Skoog, M. M. Machulda, W. K. Kremers, D. S. Knopman, C. Jack Jr, *et al.*, "Plasma and CSF neurofilament light: relation to longitudinal neuroimaging and cognitive measures," *Neurology*, vol. 93, no. 3, pp. e252–e260, 2019.

[60] H. Braak and E. Braak, "Neuropathological stageing of Alzheimer-related changes," *Acta neuropathologica*, vol. 82, no. 4, pp. 239–259, 1991.

[61] T. Schäfer and C. Ecker, "fsbrain: an R package for the visualization of structural neuroimaging data," *Biorxiv*, pp. 2020–09, 2020.