ENHANCING CLINICAL STRATIFICATION OF NEURODEGENERATION WITH VNN-INFORMED ANATOMIC FEATURES

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

[†] University at Albany, Albany, NY.
[*] University of Rochester, Rochester, NY.
[‡] University of Pennsylvania, Philadelphia, PA.

ABSTRACT

Understanding clinical heterogeneity in neurodegeneration is paramount for personalized/precision medicine. A foundational approach to identify clinically heterogeneous subtypes within neurodegenerative conditions is to perform cluster analysis of neuroimaging data that capture atrophy patterns due to neurodegeneration. However, raw anatomical features, in their own right, provide an incomplete understanding of neurodegeneration because they do not include two key factors that statistically characterize neurodegeneration: graph-theoretic information and aging. Thus, standard clustering approaches do not necessarily guarantee clinically heterogeneous clusters. Under this motivation, we seek to enrich the raw anatomic features with graph-theoretic information and aging with a principled coVariance neural network (VNN)-driven approach. In this paper, we leverage VNN model as a supervised autoencoder that achieves principled integration of the information inherent within brain morphometric features, the anatomical covariance matrix, and aging to provide enriched anatomic features, which are referred to as *VNN-informed features*. Our experiments revealed that *VNN-informed features* consistently provided a more statistically significant stratification of various neurodegenerative conditions in terms of clinical markers of disease severity than was possible through raw brain morphometric features or other commonly adopted alternatives in the literature.

***Index Terms*—** One, two, three, four, five

1. INTRODUCTION

Neurodegeneration is the progressive loss of structure or function of neurons and can be encoded within a combination of various data modalities, including neuroimaging [1], biological markers [2], and clinical readings [3, 4]. Importantly, neurodegeneration is a statistically complex phenomenon, as it is a characteristic of the healthy aging process and various neurological disorders [5, 6] with well-documented heterogeneity in factors, such as disease onset, disease severity, anatomic signatures, and disease progression. There has been an increased focus on adopting data-driven approaches to stratify various neurodegenerative conditions, such as Alzheimer's disease [7, 8], Parkinson's disease [9, 10], progressive supranuclear palsy [11], and frontotemporal dementia [12], with a broader goal of developing targeted disease-modifying therapies [13–15].

The most prevalent approach to stratify neurodegenerative conditions involves *unsupervised clustering* of relevant data modalities from patient cohorts (such as neuroimaging, biological measurements, demographic information) [16–18], with neuroimaging being the most commonly used data modality [18]. In this paper, we limit our focus to the stratification of neurodegenerative conditions using brain morphometric features. Brain morphometric features (such as cortical thickness, area, volume) are extracted from structural MRI, which is by far the most commonly adopted neuroimaging modality in clinical workflows. Stratifying neurodegenerative conditions is well-motivated for various reasons relevant to precision medicine, including varying rates of progression of the disease across different subgroups (for instance, due to certain genetic risk factors in Alzheimer's disease [19]) and tailored or adaptive interventions for specific disease subgroups [20] being the prominent ones. To clarify this aspect further, we refer the reader to Table 1 in the review article in [7]. Therein, the clinical relevance of identified clusters from neuroimaging data is substantiated by their characterization in terms of factors such as cognitive decline, executive function decline, memory decline, genetic profiles, language deficits, and polygenic risk scores. Clearly, distinct clinical and neurobiological characterizations of data-driven subtypes via post-hoc analyses are essential for the successful translation of disease stratifications to precision medicine [21, 22].

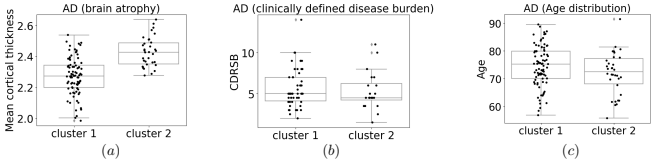

Fig. 1. Clustering AD according to brain atrophy identifies clusters significantly distinct in terms of brain atrophy, but not disease burden (in terms of CDRSB) or age distribution.

However, clustering of neuroimaging features *does not guarantee clinical heterogeneity* between identified clusters. Figure 1 provides compelling evidence in this regard. In this experiment, we leveraged hierarchical clustering of the cortical thickness features of a population of 118 individuals diagnosed with Alzheimer's disease (AD) from the well-known ADNI study [23]. Reduction in cortical thickness, or cortical thinning, is a metric of brain atrophy, with neurodegenerative conditions like AD exhibiting accelerated brain atrophy as the disease progresses. Although the two identified clusters were significantly distinct in terms of brain atrophy (Fig. 1a), they were not separated in terms of disease burden as measured by the clinical dementia rating-sum of boxes (CDRSB) scores.

CDRSB is a prevalent metric of disease burden used in clinical settings [24]. Similar observations have been reported before, for example in [25], where no significant differences in various clinical metrics were reported at the baseline across different subgroups identified using clustering of neuroimaging data of a similar cohort. We contend that clustering cortical thickness features directly is a *suboptimal* processing of the information inherent within because of two reasons: (i) the graph-theoretic information inherent within neuroimaging data [26] is not leveraged, and (ii) brain atrophy manifests in both healthy aging and neurodegeneration, but the differences between their evolution for a person of specific age to exhibit certain level of atrophy are not taken into consideration. In this context, our aim in this paper can be articulated as follows.

Aim: to propose a principled strategy for efficient processing of information inherent within brain morphometric features, such that, clustering approaches achieve better clinical stratification of neurodegeneration.

It is worth noting that the study in [25] extended their experiments to longitudinal analysis of clinical metrics, where one cluster exhibited markedly faster cognitive decline over time. However, to the best of our knowledge, the relationship between principled information processing of neuroimaging features and clinical stratification of neurodegeneration has not been explored systematically before. This relationship is crucial for the stratification of neurodegeneration in cross-sectional datasets.

The *key contribution* of this paper lies in providing a principled approach of leveraging reconstructed anatomic features derived from *coVariance neural network (VNN)-driven (supervised) autoencoder* to achieve better clinical stratification than that possible by standard anatomic features extracted from structural MRI. VNNs have recently been studied as graph neural networks (GNNs) operating on the sample covariance matrix graph, to advance the theoretical and empirical principles of statistical inference [27–32].

2. A VNN-DRIVEN APPROACH TO TRANSFORMING ANATOMIC FEATURES

Preliminaries. The anatomic features are derived from structural MRI, with each element representing a statistic (such as cortical thickness) associated with a distinct brain region. Moreover, the anatomical covariance matrix provides the graph representation of the inter-relationships between different anatomic features across the whole brain [33]. Anatomic features and anatomic covariance matrix hold significant relevance in computational neuroscience, where recent works on morphometric similarity networks have generalized the concept of anatomical covariance to include multiple modalities of information inherent within structural MRI [34] and demonstrated their relevance to identifying biomarkers [35].

To set up the VNN-driven approach technically, consider a dataset consisting of n individuals, whose anatomic features are represented by m-dimensional vectors, such that, the vector of anatomic features for individual i is given by $\mathbf{x}_i \in \mathbb{R}^{m\times 1}$. The *anatomic* covariance matrix for this dataset is estimated as

$$\mathbf{C} \triangleq \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^{\mathsf{T}}\ , \tag{1}$$

where $\bar{\mathbf{x}}$ is the sample mean of anatomic features across the dataset. Similar to GNNs that leverage *linear-shift-and-sum* operators over matrix representation of a graph as graph filters [36–38], the convolution operation in a VNN is modeled by a *coVariance filter*, given by $\mathbf{H}(\mathbf{C}) \triangleq \sum_{k=0}^{K} h_k\mathbf{C}^k$, such that, output $\mathbf{z} = \mathbf{H}(\mathbf{C})\mathbf{x}$ for input $\mathbf{x} \in \mathbb{R}^{m\times 1}$. The scalar parameters $\{h_k\}_{k=0}^{K}$ are the *filter taps* that are learned from the data. Note that the application of coVariance filter preserves the shape of $\mathbf{x}$ at the output $\mathbf{z}$.

coVariance filters and PCA. The foundational works on VNNs have leveraged the eigendecomposition of the covariance matrix $\mathbf{C}$ to establish that coVariance filter is fundamentally similar to the well-known principal component analysis (PCA)-transform [27, 30]. Specifically, given the eigendecomposition $\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathsf{T}}$, where $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{C}$ and $\mathbf{\Lambda}$ is a diagnonal matrix of eigenvalues $\{\lambda_i\}_{i=1}^{m}$ ordered as $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_m$, it can be readily checked that

$$\mathbf{V}^{\mathsf{T}}\mathbf{z} = h(\Lambda)\mathbf{V}^{\mathsf{T}}\mathbf{x} \quad \text{where} \quad h(\Lambda) = \sum_{k=0}^{K} h_k\Lambda^k$$

Thus, the *learnable* function $h(\Lambda)$ determines the contribution of specific eigenvectors within $\mathbf{V}$ to the reconstruction $\mathbf{z}$. This observation provides the foundation to further develop a coVariance filter as an autoencoder that can strategically generate representations amenable to the task at hand.

coVariance filters as supervised autoencoders. While PCA is most prominently associated with dimensionality reduction in unsupervised learning, it can also be interpreted as a *linear autoencoder* that achieves reconstruction of the input by leveraging the principal subspace of the dataset at hand (see Chapter 14 in [39] for specific analytical arguments). The coVariance filter provides a supervised learning setup in this context, where the function $h(\Lambda)$ denotes a learned transformation of the principal subspace of the covariance matrix to generate $\mathbf{z}$.

Theorem 2.1. *Consider a labeled dataset $\{\mathbf{x}_i, y_i\}_{i=1}^{n}$. The coVariance filter $\mathbf{H}(\mathbf{C})$ can be set up as a supervised autoencoder when its filter taps $\{h_k\}_{k=0}^{K}$ are fine-tuned to form an estimate $\hat{y}_i$ of label y_i from input $\mathbf{x}_i$ as follows:*

$$\hat{y}_i = \frac{1}{m}\sum_{j=1}^{m}[\mathbf{z_i}]_j, \ \ \textit{where} \ \ \mathbf{z}_i = \mathbf{H}(\mathbf{C})\mathbf{x}_i\ ,$$

and $\mathbf{z}_i$ provides the reconstruction from $\mathbf{x}_i$ fine-tuned to the learning objective of predicting y_i.

As per Theorem 2.1, the output $\mathbf{z}_i$ is the reconstruction of the input $\mathbf{x}_i$, such that, $\mathbf{z}_i = \mathbf{V}h(\mathbf{\Lambda})\mathbf{V}^{\mathsf{T}}\mathbf{x}_i$, where $h(\mathbf{\Lambda})$ determines how the principal components were exploited to achieve the reconstruction. In this paper, we focus on datasets with scalar labels $\{y_i\}$ and hence, Theorem 2.1 is specialized to this setting. Extension of Theorem 2.1 to non-scalar labels is straightforward. A coVariance filter holds a significant advantage over PCA in terms of *stability* to stochastic perturbations in the covariance matrix [27], which is necessary to ensure that the reconstruction $\mathbf{z}_i$ is not prone to small perturbations to the dataset (for instance, due to removal or addition of samples).

VNN-informed anatomic features coVariance filters are linear models with a limited learning capacity. A *perceptron* of a VNN, which is capable of learning non-linear patterns, is formed by concatenating the coVariance filter with a pointwise non-linear activation function $\sigma(\cdot)$ (e.g., $\mathsf{ReLU}, \tanh$) that

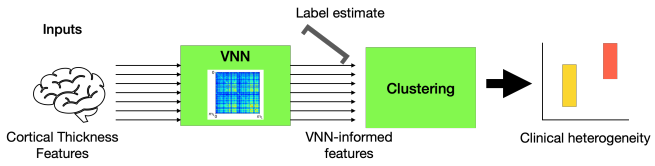

Fig. 2. Workflow for generating VNN-informed anatomic features from cortical thickness features and identifying clinical heterogeneity within a neurodegenerative condition.

satisfies $\sigma(\mathbf{u}) = [\sigma(u_1), \ldots, \sigma(u_m)]$ for $\mathbf{u} = [u_1, \ldots, u_m]$. Thus, the output of a single-layer VNN for an input $\mathbf{x}$ is $\mathbf{z} = \sigma(\mathbf{H}(\mathbf{C})\mathbf{x})$. Similar to other deep learning architectures, the capacity of VNN can be improved by adding more layers and channels for information processing [27]. We use the notation $\Phi(\mathbf{x}; \mathbf{C}, \mathcal{H})$ to compactly represent the output at the final layer of the VNN for the given input $\mathbf{x}$. The notation $\mathcal{H}$ denotes the set of all learnable parameters (filter taps) within the VNN.

Pertinent to the application at hand, the VNN model can be designed to generate the output $\Phi(\mathbf{x}; \mathbf{C}, \mathcal{H})$ with a similar dimensionality as the input $\mathbf{x}$. The learnable parameters $\mathcal{H}$ are fine-tuned to form an estimate $\hat{y}$ of a label y, such that,

$$\hat{y} = \frac{1}{m} \sum_{j=1}^{m} [\Phi(\mathbf{x}; \mathbf{C}, \mathcal{H})]_j \ .$$

Thus, given input anatomic features $\mathbf{x}$, the VNN model provides the output $\Phi(\mathbf{x}; \mathbf{C}, \mathcal{H})$, which is the anatomic representation generated by VNN as it can be projected on to the brain surface in a fashion similar to the input $\mathbf{x}$. Based on our discussion on coVariance filters leading up to their setup as supervised autoencoders in Theorem 2.1 and the fact that coVariance filters form the elementary information processing module within a VNN, we argue that the VNN model forms the output $\Phi(\mathbf{x}; \mathbf{C}, \mathcal{H})$ from input $\mathbf{x}$, at least in part, by exploiting the principal components of the covariance matrix $\mathbf{C}$. However, unlike the case of the coVariance filter in this context, this argument may not be readily provable for a generic VNN model. Hence, we articulate our argument that VNN can be interpreted as a supervised autoencoder in the following proposition.

Proposition 2.2. *Consider a labeled dataset $\{\mathbf{x}_i, y_i\}_{i=1}^n$. The VNN model can be set up as a supervised autoencoder when its learnable parameters $\mathcal{H}$ are fine-tuned to form an estimate $\hat{y}_i$ of label y_i from input $\mathbf{x}_i$ as follows:*

$$\hat{y}_i = \frac{1}{m} \sum_{j=1}^{m} [\Phi(\mathbf{x}_i; \mathbf{C}, \mathcal{H})]_j,$$

and $\Phi(\mathbf{x}_i; \mathbf{C}, \mathcal{H})$ provides the reconstruction from $\mathbf{x}_i$ fine-tuned to the learning objective of predicting y_i.

As per Proposition 2.2, the output $\Phi(\mathbf{x}_i; \mathbf{C}, \mathcal{H})$ is the reconstruction of the input $\mathbf{x}_i$, achieved at least in part by $\mathcal{H}$-mediated exploitation of the principal components. We refer to the output $\Phi(\mathbf{x}_i; \mathbf{C}, \mathcal{H})$ as the *VNN-informed* feature generated from anatomic feature $\mathbf{x}_i$. Based on the motivation described in 'Using age as the label' below, we set the label y_i to be chronological age.

Using age as the label. In this paper, we select 'chronological age' as the label to train a VNN. Age is one of the most relevant biological variables in statistical analyses of neurodegeneration with neuroimaging because of two key reasons: (i) age is a leading risk factor for various neurodegenerative conditions [40]; and (ii) brain anatomic features evolve with healthy aging and exhibit accelerated or anomalous behaviors for various neurodegenerative conditions relative to the healthy population [41].

3. STRATIFYING NEURODEGENERATION WITH VNN-INFORMED FEATURES

The VNN-driven approach to transforming anatomic features in Section 2 provides a principled way to combine the information inherent within the raw anatomic features, anatomic covariance matrix, and age. In this section, we describe the experiment setup to stratify a neurodegenerative cohort using VNN-informed features.

Pre-training a VNN model on healthy cohort. We used the healthy control population from the OASIS-3 dataset [42] to pre-train the VNN model. The cortical thickness data consists of $m = 68$ features across the cortex, curated according to the Desikan-Killiany brain atlas [43]. The VNN comprised 2 layers of width 61, totaling $22,570$ learnable parameters. The unweighted mean of the output at the final layer formed the age estimate. We hypothesized the statistical properties of VNN-informed features derived from this pre-VNN to be different and more informative than their raw cortical thickness counterparts, due to the inherent convolution over the anatomic covariance matrix and the aging context infused during the VNN pre-training.

Stratifying neurodegenerative conditions with VNN-informed features derived from the pre-trained VNN. For a given cohort with neurodegenerative conditions, we clustered the VNN-informed features (after *mean normalization*) using hierarchical clustering [44] and non-negative matrix factorization (NMF) [45]. These two clustering algorithms have been previously adopted in the literature [46–51]. The primary goal of this set of experiments was to determine whether clustering VNN-informed features led to more *clinically heterogeneous* clusters relative to raw cortical thickness features or other commonly used baselines in the disease subtyping literature. Thus, the metric of interest was determined in terms of the group level differences (ANCOVA with age as covariate) between the distributions of CDRSB or NP3TOT scores (see Section 4.1) for identified clusters in the disease cohorts. We prescribed the number of clusters to be 2 for all clustering experiments. The rationale behind this choice is that the datasets leveraged are moderate in size, and the choice of 2 clusters provides the maximal statistical power and a more convincing picture to gauge clinical heterogeneity within the sub-populations than that feasible in scenarios with more than 2 groups (due to possibility of extremely small subgroups, for example).

Since the weights of the pre-trained VNN model were not altered, it was oblivious to the identity of the neurodegenerative condition being stratified. For hierarchical clustering, dendrograms were constructed for VNN-informed features derived for neurodegenerative conditions based on the nearest-neighbor chain method and Euclidean distance (see ??).

Condition (clinical marker)	F-value (p-value) for ANCOVA with Age as covariate			
	Raw CT	[CT, Age]	Residualized CT	VNN-informed features
AD (CDRSB)	0.46 (0.498)	0.16 (0.68)	0.65 (0.42)	**4.46 (0.038)**
bvFTD (CDRSB)	6.81 (0.0012)	1.47 (0.23)	**7.4 (0.009)**	6.13 (0.016)
svPPA (CDRSB)	0.88 (0.355)	0.65 (0.42)	0.88 (0.355)	**4.75 (0.037)**
PNFA (CDRSB)	–	0.009 (0.92)	–	**14.01 (9.1×10^{-4})**
PD (NP3TOT)	10.78 (1.1×10^{-3})	4.48 (0.028)	0.096 (0.757)	**14.08 (2×10^{-4})**

Table 1. Results for clusters derived using hierarchical clustering.

4. RESULTS

4.1. Datasets and clinical markers of disease severity

We studied five neurodegenerative conditions: Alzheimer's disease (AD), Parkinson's disease (PD), Behavioral Frontotemporal Degeneration (bvFTD), Semantic Variant Primary Progressive Aphasia (svPPA), and Progressive Non-fluent Aphasia (PNFA). All datasets are described in Appendix **??**. For all conditions, we considered cortical thickness features that had been derived from T1-weighted MRI images acquired on 3.0 Tesla MRI scanners. The common theme across all considered neurodegenerative conditions studied is that (i) they are characterized by cortical atrophy and (ii) age is a prominent risk factor or determinant of disease onset and severity [52–55].

We seek to compare VNN-informed features against various baselines that involve clustering of raw cortical thickness features accompanied with age information. To carry out these comparisons, we consider the following prominent clinical markers of disease severity.

CDRSB. We considered Clinical Dementia Rating Sum-of-boxes (CDRSB) as the metric of disease severity in AD, bvFTD, svPPA, and PNFA subgroups.

NP3TOT score. The MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) is a comprehensive assessment designed to monitor the burden and extent of PD [56]. Part 3 of MDS-UPDRS is focused on the motor examination, and clinicians employ this score in daily clinical practice to objectively track motor performance and progression over time [57]. Here, we use the notation NP3TOT to refer to the score associated with Part 3 of MDS-UPDRS.

4.2. Stratification of Neurodegenerative Conditions

We obtained the two clusters from VNN-informed anatomical features using hierarchical and NMF-based clustering algorithms for all neurodegenerative conditions studied. The anatomic characterizations of these clusters as determined by VNN-informed features will be discussed subsequently. First, we focus on the comparisons against different baselines in terms of clinical stratifications achieved.

Anatomical baselines. We first considered three *anatomic features-driven baselines* that have been widely adopted in the existing studies on subtyping neurodegeneration [7, 58–60]:

- *Raw CT*: Raw cortical thickness features;
- *[CT, Age]*: Raw cortical thickness features concatenated with age; and
- *Residualized CT*: Age-corrected raw cortical thickness features. For every cortical region, this correction was performed by a linear regression model that had been trained on the respective healthy cohort (HC) to predict chronological age from the associated cortical thickness data. This pre-trained linear regression model was then applied on the data with the neurodegenerative cohort, and the residual of the model formed the residualized CT statistic as the baseline.

Our objective was to compare the *clinical* stratification of identified clusters using all approaches. Age was used as a covariate variable to avoid reporting of trivial clustering according to the age of the patient population. **Table 1** tabulates the results derived using agglomerative hierarchical clustering with Euclidean distance metric. The approaches that yielded the largest clinical heterogeneity have been emphasized.

The results in Table 1 convey that the clusters identified using VNN-informed features were the most clinically stratified in the AD, svPPA, PNFA, and PD disease cohorts. The entries corresponding to *Raw CT* and *Residualized CT* for the PNFA cohort in Table 1 are missing as hierarchical clustering could not identify a second cluster of significant size (larger than 10% of the total cohort size).

Interestingly, none of the three baselines could identify clusters in the AD, svPPA, and PNFA cohorts with significantly different CDRSB score. However, the clusters identified using VNN-informed features in the AD cohort differed significantly in terms of CDRSB scores (ANCOVA with age as covariate, p-value < 0.05). Moreover, for bvFTD, the clusters identified using VNN-informed features significantly differed in terms of CDRSB score, but were outperformed by clusters derived using *Raw CT* and *Residualized CT* baselines.

4.3. Additional Experiments

5. DISCUSSION

We have presented an innovative VNN-driven pipeline that achieves a principled amalgamation of anatomic features, age information, and covariance to yield VNN-informed anatomic features. The key contributions of this paper are both conceptual and empirical. We considered 5 neurodegenerative conditions: 4 within the family of Alzheimer's disease and related dementias, plus Parkinson's disease. VNN-informed features consistently achieved more statistically significant stratification in terms of markers of disease burden (cognitive impairment and motor impairment) relative to raw cortical thickness features, or, age-corrected cortical thickness features. The strength of the results lies in the consistency of findings across 4 out of 5 different populations.

PCA-driven baselines. We also considered two *PCA-driven baselines* as PCA approach holds the most relevance with respect to VNNs. These approaches reduced the dimen-

sionality of the neuroimaging datasets from neurodegenerative cohorts according to different sets of principal components.

- *PCA Approach 1:* In this approach, we leveraged the principal components of the anatomic covariance matrix estimated from the healthy cohorts.

- *PCA Approach 2:* In this approach, we leveraged the principal components of the anatomic covariance matrix estimated from the neurodegenerative disease cohorts.

We clarify that we did not have a pre-defined hypothesis to select the number of principal components for these approaches. Hence, we investigated the performances for these approaches (in terms of F-values for ANCOVA over clinical scores across 2 clusters identified via hierarchical clustering in the disease groups) averaged over the clusters identified based on the reduced datasets according to the first K number of principal components (K varying from 1 to 15) derived from the features of respective disease populations. Both baselines achieved less significant clinical stratification in the two clusters (see Appendix **??** for specific results). These results also highlighted the value of using age as the label for VNN-driven approach.

Here, we also highlight that the performances achieved by both PCA-driven approaches were not robust to the choice of principal components used. As an example, the clusters identified from PCA Approach 1 with the first 9 or 11 principal components exhibited much smaller clinical heterogeneity (F-values 0.795 and 1.70) relative to the clusters identified using PCA-reduced features from the first 10 principal components (F-value = 8.92). This observation attests to the well-recognized lack of reproducibility of PCA-driven statistical approaches, which is overcome within the seminal works on VNNs (Sihag et al, 2022).

Anatomic characterization. Next, we report the anatomic characterizations of the clusters identified using VNN-informed features for all neurodegenerative conditions. In the main paper, we focus only on the clusters identified using hierarchical clustering. For conciseness, we report the anatomic characterizations of two clusters on the same brain surface by performing element-wise group difference analyses of VNN-informed features derived from cluster 1 and cluster 2, for all neurodegenerative conditions.

Figure 3 illustrates the anatomic patterns associated with the results of group-level difference analysis from ANOVA between cluster 1 and cluster 2 (identified using hierarchical clustering), for different neurodegenerative conditions. Each sub-figure in Fig. 3 has been obtained by projecting the F-value of ANOVA for element-wise group level comparison of VNN-informed features of cluster 1 and cluster 2. Moreover, in Fig. 3, green color on the brain surface represents the directionality of the mean associated element in VNN-informed features across cluster 1 being larger than that across cluster 2, and red color encodes the opposite. We uniformly associated 'Cluster 2' with the subgroup with higher disease burden as determined after clustering of VNN-informed features. In this context, the brain regions marked with red color in Fig. 3 indeed correlated with the regions known to be relevant for the respective diseases (see additional discussions in Appendix **??**).

Explaining anatomic characterization with principal components. We further explored whether there were differences in how the pre-trained VNN model processed the raw cortical thickness features for the identified clusters as per the *inner product metrics* in Section 2. To this end, we evaluated and reported the inner product of normalized VNN-informed features (norm = 1) and the eigenvectors of the respective anatomical covariance matrix. For the AD cohort, the results in Fig. **??** depict the inner product metrics between the VNN-informed features and the first 20 eigenvectors of the respective anatomic covariance matrix. Notably, certain eigenvectors were more strongly linked to VNN-informed features than others. Interestingly, the 0-th, first, and 18-th eigenvectors exhibited significantly different (ANOVA, p-value after Bonferroni correction < 0.05) inner product metrics for the clusters in AD identified using hierarchical clustering. Similar observations for other disease cohorts have been reported in Appendix **??**. The results in this context corroborated that the pre-trained VNN model indeed leveraged the principal components of the covariance difference differently to yield distinct clusters.

6. DISCUSSION

We have presented an innovative VNN-driven pipeline that achieves a principled amalgamation of anatomic features, age information, and covariance to yield VNN-informed anatomic features. The key contributions of this paper are both conceptual and empirical and have been highlighted below.

Enhanced clinical stratification of neurodegenerative conditions. We considered 5 neurodegenerative conditions: 4 within the family of Alzheimer's disease and related dementias, plus Parkinson's disease. VNN-informed features consistently achieved more statistically significant stratification in terms of markers of disease burden (cognitive impairment and motor impairment) relative to raw cortical thickness features, or, age-corrected cortical thickness features. The strength of the results lies in the consistency of findings across 5 different populations.

Transparent construction of VNN-informed features. Building upon the interpretation of the VNN as an autoencoder (Theorem 2.1 and Proposition 2.2), our experiments have revealed that the identified clusters from VNN-informed features could be tied to differences in how the principal components of the covariance matrix were leveraged by the VNN to construct the associated VNN-informed features. Such transparency is critical for applications in healthcare, as it provides a potential mechanism for the intended end-users (clinicians and patients) to trust the findings yielded by the VNN model.

Broadly, the contributions in this paper lie at the intersection of machine learning, computational neuroscience, and precision medicine. The innovative methods proposed in this paper could potentially open up new perspectives to leveraging anatomic features in clinical workflows and analysis of heterogeneity within neurodegenerative conditions in precision medicine.

Limitations. The impact of our results is limited by the size of the different datasets considered. This concern is partly mitigated by the fact that all datasets considered are publicly available and hence, the findings are repeatable by independent researchers without significant hurdles. Nevertheless, our experiments may not have captured the true heterogeneity within the neurodegenerative conditions due to limited data size and consideration of only two clusters.

Our study is methodology-focused and hence, provides limited impact for clinical audiences. Specifically, we modeled

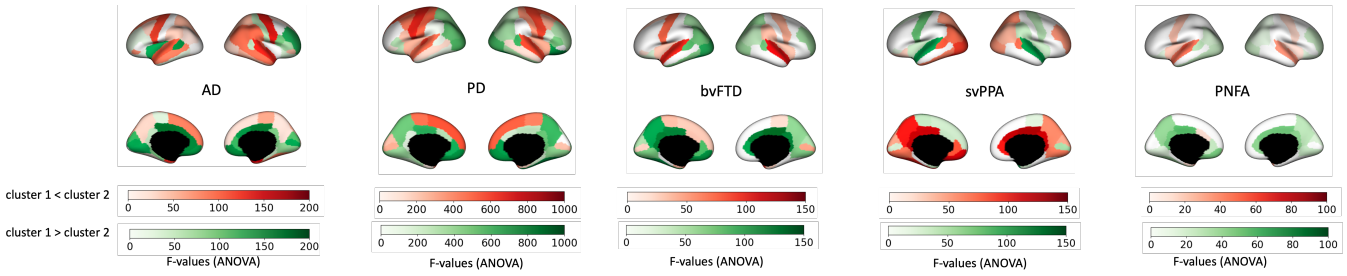

Fig. 3. Anatomic characterization of the group level differences (ANOVA) between VNN-informed features from clusters 1 and 2 (hierarchical clustering), for different neurodegenerative conditions. F-values for ANOVA with p-value < 0.05 after Bonferroni correction for multiple comparisons have been plotted on the brain surfaces. Green and red colors encode the directionality of group difference.

all experiments within the narrow regime defined by cortical thickness features and two specific measures of disease burden. Although our methodology is innovative and our experiments comprehensive, the reported results may hold little relevance for clinical use. Hence, a more comprehensive study with a broader set of anatomic features, neurobiological factors, and clinical markers must be conducted to identify the true clinical impact.

The method of constructing VNN-informed features is restricted to the 'Age' factor. In general, age may not have a uniform impact across all pathologies, which potentially explains why raw cortical thickness features achieved comparable clinical stratification in bvFTD (Table 1). Hence, the neuropathology of diseases must be taken into account in a comprehensive clinical study in the future. Moreover, other factors, such as sex, could be considered for construction of VNN-informed features.

We have not compared our results against other deep learning based autoencoders, which could potentially provide similar functionality. Hence, we cannot claim that VNNs achieve the best clinical stratification possible within the considered setting. However, it's notable that the findings reported by the VNN-driven method are backed by sound statistical analysis and technical arguments towards explainability or transparency. Such functionality is not universal across deep learning-based autoencoders. This paper potentially opens up a new application domain, where the effectiveness of other deep learning approaches could be compared in future studies.

7. REFERENCES

[1] Shannon L Risacher and Andrew J Saykin, "Neuroimaging biomarkers of neurodegenerative diseases and dementia," in *Seminars in neurology*. Thieme Medical Publishers, 2013, vol. 33, pp. 386–416.

[2] Leslie M Shaw, Magdalena Korecka, Christopher M Clark, Virginia M-Y Lee, and John Q Trojanowski, "Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics," *Nature reviews Drug discovery*, vol. 6, no. 4, pp. 295–303, 2007.

[3] Michele Terzaghi, Elena Sinforiani, Chiara Zucchella, Elena Zambrelli, Chiara Pasotti, Valter Rustioni, and Raffaele Manni, "Cognitive performance in rem sleep behaviour disorder: a possible early marker of neurodegenerative disease?," *Sleep medicine*, vol. 9, no. 4, pp. 343–351, 2008.

[4] Tim J Anderson and Michael R MacAskill, "Eye movements in patients with neurodegenerative disorders," *Nature Reviews Neurology*, vol. 9, no. 2, pp. 74–85, 2013.

[5] Serge Przedborski, Miquel Vila, Vernice Jackson-Lewis, et al., "Series introduction: Neurodegeneration: What is it and where are we?," *The Journal of clinical investigation*, vol. 111, no. 1, pp. 3–10, 2003.

[6] Dietmar Rudolf Thal, Kelly Del Tredici, and Heiko Braak, "Neurodegeneration in normal brain aging and disease," *Science of aging knowledge environment*, vol. 2004, no. 23, pp. pe26–pe26, 2004.

[7] Mohamad Habes, Michel J Grothe, Birkan Tunc, Corey McMillan, David A Wolk, and Christos Davatzikos, "Disentangling heterogeneity in alzheimer's disease and related dementias using data-driven methods," *Biological psychiatry*, vol. 88, no. 1, pp. 70–82, 2020.

[8] Fedor Levin, Daniel Ferreira, Catharina Lange, Martin Dyrba, Eric Westman, Ralph Buchert, Stefan J Teipel, Michel J Grothe, and Alzheimer's Disease Neuroimaging Initiative, "Data-driven fdg-pet subtypes of alzheimer's disease-related neurodegeneration," *Alzheimer's research & therapy*, vol. 13, pp. 1–14, 2021.

[9] Mark Frasier, Brian K Fiske, and Todd B Sherer, "Precision medicine for parkinson's disease: The subtyping challenge," *Frontiers in aging neuroscience*, vol. 14, pp. 1064057, 2022.

[10] Matthew Brendel, Chang Su, Yu Hou, Claire Henchcliffe, and Fei Wang, "Comprehensive subtyping of parkinson's disease patients with similarity fusion: A case study with biofind data," *npj Parkinson's Disease*, vol. 7, no. 1, pp. 83, 2021.

[11] William J Scotton, Cameron Shand, Emily Todd, Martina Bocchetta, David M Cash, Lawren VandeVrede, Hilary Heuer, Alexandra L Young, Neil Oxtoby, et al., "Uncovering spatiotemporal patterns of atrophy in progressive supranuclear palsy using unsupervised machine learning," *Brain communications*, vol. 5, no. 2, pp. fcad048, 2023.

[12] Emma L Van Der Ende, Esther E Bron, Jackie M Poos, Lize C Jiskoot, Jessica L Panman, Janne M Papma, Lieke H Meeter, Elise GP Dopper, Carlo Wilke, Matthis Synofzik, et al., "A data-driven disease progression model of fluid biomarkers in genetic frontotemporal dementia," *Brain*, vol. 145, no. 5, pp. 1805–1817, 2022.

[13] Alberto J Espay, "Models of precision medicine for neurodegeneration," *Handbook of Clinical Neurology*, vol. 192, pp. 21–34, 2023.

[14] Harald Hampel, Andrea Vergallo, George Perry, Simone Lista, Alzheimer Precision Medicine Initiative, et al., "The alzheimer precision medicine initiative," *Journal of Alzheimer's Disease*, vol. 68, no. 1, pp. 1–24, 2019.

[15] David J Irwin, Nigel J Cairns, Murray Grossman, Corey T McMillan, Edward B Lee, Vivianna M Van Deerlin, Virginia M-Y Lee, and John Q Trojanowski, "Frontotemporal lobar degeneration: defining phenotypic diversity through personalized medicine," *Acta neuropathologica*, vol. 129, pp. 469–491, 2015.

[16] Jennifer L Whitwell, Scott A Przybelski, Stephen D Weigand, Robert J Ivnik, Prashanthi Vemuri, Jeffrey L Gunter, Matthew L Senjem, Maria M Shiung, Bradley F Boeve, David S Knopman, et al., "Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study," *Brain*, vol. 132, no. 11, pp. 2932–2946, 2009.

[17] Nonie Alexander, Daniel C Alexander, Frederik Barkhof, and Spiros Denaxas, "Identifying and evaluating clinical subtypes of alzheimer's disease in care electronic health records using unsupervised machine learning," *BMC Medical Informatics and Decision Making*, vol. 21, pp. 1–13, 2021.

[18] Pindong Chen, Shirui Zhang, Kun Zhao, Xiaopeng Kang, Timothy Rittman, and Yong Liu, "Robustly uncovering the heterogeneity of neurodegenerative disease by using data-driven subtyping in neuroimaging: a review," *Brain Research*, vol. 1823, pp. 148675, 2024.

[19] Lars Bertram and Rudolph E Tanzi, "The genetics of alzheimer's disease," *Progress in molecular biology and translational science*, vol. 107, pp. 79–100, 2012.

[20] Steven E Arnold, Bradley T Hyman, Rebecca A Betensky, and Hiroko H Dodge, "Pathways to personalized medicine—embracing heterogeneity for progress in clinical therapeutics research in alzheimer's disease," *Alzheimer's & Dementia*, vol. 20, no. 10, pp. 7384–7394, 2024.

[21] Connie Marras, Seyed-Mohammad Fereshtehnejad, Daniela Berg, Nicolaas I Bohnen, Kathy Dujardin, Roberto Erro, Alberto J Espay, Glenda Halliday, Jacobus J Van Hilten, Michele T Hu, et al., "Transitioning from subtyping to precision medicine in parkinson's disease: A purpose-driven approach," *Movement Disorders*, vol. 39, no. 3, pp. 462–471, 2024.

[22] Serena Verdi, Andre F Marquand, Jonathan M Schott, and James H Cole, "Beyond the average patient: how neuroimaging models can address heterogeneity in dementia," *Brain*, vol. 144, no. 10, pp. 2946–2953, 2021.

[23] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al., "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.

[24] John C Morris, "The clinical dementia rating (cdr): current version and scoring rules.," *Neurology*, 1993.

[25] Shannon L Risacher, Wesley H Anderson, Arnaud Charil, Peter F Castelluccio, Sergey Shcherbinin, Andrew J Saykin, Adam J Schwarz, Alzheimer's Disease Neuroimaging Initiative, and Alzheimer's Disease Neuroimaging Initiative, "Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline," *Neurology*, vol. 89, no. 21, pp. 2176–2186, 2017.

[26] Olaf Sporns, "Graph theory methods: applications in brain networks," *Dialogues in clinical neuroscience*, vol. 20, no. 2, pp. 111–121, 2018.

[27] Saurabh Sihag, Gonzalo Mateos, Corey McMillan, and Alejandro Ribeiro, "coVariance neural networks," in *Proc. Conference on Neural Information Processing Systems*, Nov. 2022.

[28] Andrea Cavallo, Madeline Navarro, Santiago Segarra, and Elvin Isufi, "Fair covariance neural networks," *arXiv preprint arXiv:2409.08558*, 2024.

[29] Andrea Cavallo, Zhan Gao, and Elvin Isufi, "Sparse covariance neural networks," *arXiv:2410.01669*, vol. cs.LG, 2024.

[30] Andrea Cavallo, Mohammad Sabbaqi, and Elvin Isufi, "Spatiotemporal covariance neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2024, pp. 18–34.

[31] Saurabh Sihag, Gonzalo Mateos, Corey McMillan, and Alejandro Ribeiro, "Transferability of covariance neural networks," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–16, 2024.

[32] Saurabh Sihag, Gonzalo Mateos, Corey McMillan, and Alejandro Ribeiro, "Explainable brain age prediction using covariance neural networks," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

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

[34] Jakob Seidlitz, František Váša, Maxwell Shinn, Rafael Romero-Garcia, Kirstie J Whitaker, Petra E Vértes, Konrad Wagstyl, Paul Kirkpatrick Reardon, Liv Clasen, Siyuan 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.

[35] Paola Galdi, Manuel Blesa, David Q Stoye, Gemma Sullivan, Gillian J Lamb, Alan J Quigley, Michael J Thrippleton, Mark E Bastin, and James P Boardman, "Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth," *NeuroImage: Clinical*, vol. 25, pp. 102195, 2020.

[36] Antonio Ortega, Pascal Frossard, Jelena Kovačević, José MF Moura, and Pierre Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[37] Fernando Gama, Elvin Isufi, Geert Leus, and Alejandro Ribeiro, "Graphs, convolutions, and neural networks: From graph filters to graph neural networks," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 128–138, 2020.

[38] Elvin Isufi, Fernando Gama, David I Shuman, and Santiago Segarra, "Graph filters for signal processing and machine learning on graphs," *IEEE Transactions on Signal Processing*, vol. 72, pp. 4745–4781, 2024.

[39] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep learning*, MIT press, 2016.

[40] Denham Harman, "Alzheimer's disease pathogenesis: role of aging," *Annals of the New York Academy of Sciences*, vol. 1067, no. 1, pp. 454–460, 2006.

[41] M Ethan MacDonald and G Bruce Pike, "Mri of healthy brain aging: A review," *NMR in Biomedicine*, vol. 34, no. 9, pp. e4564, 2021.

[42] Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al., "OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease," *MedRxiv*, 2019.

[43] Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley 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.

[44] Fionn Murtagh and Pedro Contreras, "Algorithms for hierarchical clustering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, 2012.

[45] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[46] Carme Uribe, Barbara Segura, Hugo Cesar Baggio, Alexandra Abos, Maria Jose Marti, Francesc Valldeoriola, Yaroslau Compta, Nuria Bargallo, and Carme Junque, "Patterns of cortical thinning in nondemented parkinson's disease patients," *Movement Disorders*, vol. 31, no. 5, pp. 699–708, 2016.

[47] Carme Uribe, Barbara Segura, Hugo Cesar Baggio, Alexandra Abos, Anna Isabel Garcia-Diaz, Anna Campabadal, Maria Jose Marti, Francesc Valldeoriola, Yaroslau Compta, Eduard Tolosa, et al., "Cortical atrophy patterns in early parkinson's disease patients using hierarchical cluster analysis," *Parkinsonism & related disorders*, vol. 50, pp. 3–9, 2018.

[48] Young Noh, Seun Jeon, Jong Min Lee, Sang Won Seo, Geon Ha Kim, Hanna Cho, Byoung Seok Ye, Cindy W Yoon, Hee Jin Kim, Juhee Chin, et al., "Anatomical heterogeneity of alzheimer disease: based on cortical thickness on mris," *Neurology*, vol. 83, no. 21, pp. 1936–1944, 2014.

[49] Aristeidis Sotiras, Susan M Resnick, and Christos Davatzikos, "Finding imaging patterns of structural covariance via non-negative matrix factorization," *Neuroimage*, vol. 108, pp. 1–16, 2015.

[50] Na Lu, Tengfei Li, Jinjin Pan, Xiaodong Ren, Zuren Feng, and Hongyu Miao, "Structure constrained seminonnegative matrix factorization for eeg-based motor imagery classification," *Computers in Biology and Medicine*, vol. 60, pp. 32–39, 2015.

[51] Huan Lao and Xuejun Zhang, "Regression and classification of alzheimer's disease diagnosis using nmf-tdnet features from 3d brain mr image," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 1103–1115, 2021.

[52] Lorenzo Pini, Michela Pievani, Martina Bocchetta, Daniele Altomare, Paolo Bosco, Enrica Cavedo, Samantha Galluzzi, Moira Marizzoni, and Giovanni B Frisoni, "Brain atrophy in alzheimer's disease and aging," *Ageing research reviews*, vol. 30, pp. 25–48, 2016.

[53] Massimo Filippi, Elisabetta Sarasso, Noemi Piramide, Tanja Stojkovic, Iva Stankovic, Silvia Basaia, Andrea Fontana, Aleksandra Tomic, Vladana Markovic, Elka Stefanova, et al., "Progressive brain atrophy and clinical evolution in parkinson's disease," *NeuroImage: Clinical*, vol. 28, pp. 102374, 2020.

[54] Howard J Rosen, Maria Luisa Gorno-Tempini, WP Goldman, RJ Perry, N Schuff, Michael Weiner, R Feiwell, JH Kramer, and Bruce L Miller, "Patterns of brain atrophy in frontotemporal dementia and semantic dementia," *Neurology*, vol. 58, no. 2, pp. 198–208, 2002.

[55] Marina Boccardi, Mikko P Laakso, Lorena Bresciani, Samantha Galluzzi, Cristina Geroldi, Alberto Beltramello, Hilkka Soininen, and Giovanni B Frisoni, "The mri pattern of frontal and temporal brain atrophy in fronto-temporal dementia," *Neurobiology of aging*, vol. 24, no. 1, pp. 95–103, 2003.

[56] Christopher G Goetz, Barbara C Tilley, Stephanie R Shaftman, Glenn T Stebbins, Stanley Fahn, Pablo Martinez-Martin, Werner Poewe, Cristina Sampaio, Matthew B Stern, Richard Dodel, et al., "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[57] Joseph G Hentz, Shyamal H Mehta, Holly A Shill, Erika Driver-Dunckley, Thomas G Beach, and Charles H Adler, "Simplified conversion method for unified parkinson's disease rating scale motor examinations," *Movement Disorders*, vol. 30, no. 14, pp. 1967–1970, 2015.

[58] Serena Verdi, Seyed Mostafa Kia, Keir XX Yong, Duygu Tosun, Jonathan M Schott, Andre F Marquand, and James H Cole, "Revealing individual neuroanatomical heterogeneity in alzheimer disease using neuroanatomical normative modeling," *Neurology*, vol. 100, no. 24, pp. e2442–e2453, 2023.

[59] Konstantinos Poulakis, Joana B Pereira, J-Sebastian Muehlboeck, Lars-Olof Wahlund, Örjan Smedby, Giovanni Volpe, Colin L Masters, David Ames, Yoshiki Niimi, Takeshi Iwatsubo, et al., "Multi-cohort and longitudinal bayesian clustering study of stage and subtype in alzheimer's disease," *Nature communications*, vol. 13, no. 1, pp. 4566, 2022.

[60] Aoyan Dong, Nicolas Honnorat, Bilwaj Gaonkar, and Christos Davatzikos, "Chimera: clustering of heterogeneous disease effects via distribution matching of imaging patterns," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 612–621, 2015.