Alzheimer's Disease Classification Using Cluster-based Labelling for Graph Neural Network on Tau PET Imaging and Heterogeneous Data

Niamh McCombe[1, #], Jake Bamrah[1, #], Jose M. Sanchez-Bornot[1], David P. Finn[2], Paula L. McClean[3], KongFatt Wong-Lin[1, *], Alzheimer's Disease Neuroimaging Initiative (ADNI)

[1]Intelligent Systems Research Centre, School of Computing, Engineering and Intelligent Systems, Ulster University, Magee campus, Derry~Londonderry, Northern Ireland, UK
[2]Pharmacology and Therapeutics, Galway Neuroscience Centre, Centre for Pain Research, and School of Medicine, National University of Ireland Galway, Galway, Ireland
[3]Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute, Clinical Translational Research and Innovation Centre (C-TRIC), Ulster University, Derry~Londonderry, Northern Ireland, UK

[#]Joint first authors

*** Correspondence Author:** k.wong-lin@ulster.ac.uk

Keywords: Alzheimer's disease diagnosis, tau PET neuroimaging, graph neural network GNN, deep learning, unsupervised learning, UMAP, clustering, ADNI data

Abstract

Graphical neural networks (GNNs) offer the opportunity to provide new insights into the diagnosis, prognosis and contributing factors of Alzheimer's disease (AD). However, previous studies using GNNs on AD did not incorporate tau PET neuroimaging features or, more importantly, have applied class labels based only on clinician diagnosis, which can be inconsistent. This study addresses these limitations by applying cluster-based labelling on heterogeneous data, including tau PET neuroimaging data, before applying GNN for classification. The data comprised sociodemographic, medical/family history, cognitive and functional assessments, apolipoprotein E (APoE) ε4 allele status, and combined tau PET neuroimaging and MRI data. We identified 5 clusters embedded in a 5-dimensional nonlinear UMAP space. Further projection onto a 3-dimensional UMAP space for visualisation, and using association rule and information gain algorithms, the individual clusters were revealed to have specific feature characteristics, specifically with respect to clinical diagnosis of AD, gender, parental history of AD, age, and underlying neurological risk factors. In particular, there was one cluster comprised mainly of clinician diagnosis of AD. We re-labelled the diagnostic classes based around this AD cluster such that AD cases occurred only within this cluster while those AD cases outside of it were re-labelled as a prodromal stage of AD. A post-hoc analysis supported the re-labelling of these cases given their similar brain tau PET levels, lying between those of the remaining AD and non-AD cases. We then trained a GNN model using the re-labelled data and compared the AD classification performance to that trained using clinician diagnostic labels. We found that GNN with re-labelled data had significantly higher average classification accuracy ($p=0.011$) and less variability ($93.2\pm0.03\%$) than with clinician labelled data ($90.06\pm0.04\%$). During model training, the cluster-labelled GNN model also converged faster than that using clinician labels. Overall, our work demonstrates that more objective cluster-based labels are viable for training GNN on heterogeneous data for diagnosing AD and cognitive decline, especially when aided by tau PET neuroimaging data.

1 INTRODUCTION

Alzheimer's disease (AD) is a pernicious and progressive brain disease where neurodegeneration is commonly manifested through a variety of comorbidities such as cognitive degradation, motor impairment, speech disturbance and psychiatric changes (Burns and Iliffe, 2009). AD exists as a subset of dementia, which encompasses a large variety of neurological disorders. Dementia is one of the leading causes of disability worldwide with approximately 50 million clinically diagnosed cases globally (WHO, 2020). As of 2019, AD contributes to between 60-70% of all dementia cases with the disease predominantly affecting those aged 65 and over (Ferri et al., 2005). AD generally manifests itself in two forms, familial AD and sporadic AD, of which the latter constitutes the majority (Wong-Lin et al., 2020), and will henceforth be the focus here.

With the availability of increasingly complex and heterogeneous dementia data and still sub-optimal dementia diagnostic procedures, decision support systems, with the aid of machine learning (ML) algorithms, are gradually becoming important (Wong-Lin et al., 2020). Recently, there has been an increase in a specific ML approach to detect AD – convolutional neural networks (CNNs) (Wen et al., 2020). In particular, graph neural networks (GNNs), which are generalised versions of CNNs (e.g. graph convolutional neural networks, GCNs), allows the number of network node connections to vary and the nodes are not ordered in an Euclidean manner (Gori et al., 2005; Scarselli et al., 2009). GNNs also have the potential for interpretability (Ying et al., 2019). GNNs have been applied to a variety of biomedical or disease diagnostic applications in recent years (e.g. Parisot et al., 2018; Sun et al., 2021) due to their innate ability to handle diverse and complex graphical data (Gori et al., 2005; Scarselli et al., 2009). By using a neighbourhood of nodes (representing human participants) that comprise various disease factors or variables, participants' similarities can be derived from the relationships that exist between each node (edge weights).

With respect to predicting AD, a variety of GNN approaches have been applied. In particular, they have been applied on different types of data (Parisot et al., 2018; Kazi et al., 2019a; Kazi et al., 2019b; Wen et al., 2020; Vivar et al., 2021; Zhu and Razavian, 2021) and even single-cell data (Wang et al., 2021). A limitation of such GNN based approaches was that of graph rigidity, where the final graph structure was limited when inducting new nodes. Such an example was provided in Parisot et al. (2018) where the resulting model required a new model iteration to be trained with each new node addition. More robust GNN techniques have been proposed (Zhu and Razavian, 2021; Song et al., 2021). In particular, in Song et al., (2021), a flexible GNN was proposed to solve the fixed graph structure problem by adopting a meta-learning strategy, specifically metric-learning, which was used to infer node similarity using a trainable similarity function, facilitating the use of heterogeneous data types, including magnetic resonance imaging (MRI) data.

Complementing the supervised learning of GNN, unsupervised learning such as dimensional reduction, clustering and data visualisation have been used for providing further insights into heterogeneous AD data (e.g. Alashwal et al., 2019; Wang et al., 2021; Katabathula et al., 2021). In particular, some studies using clustering methods had identified sub-populations that were relatively homogeneous based on clinical and biological features (Gamberger et al., 2016; Martí-Juan et al., 2019; Ferreira et al., 2019; Prakash et al., 2021, Alexander et al., 2020; Alexander et al., 2021), differing rates of cognitive decline of sub-groups of prodromal AD patients (Gamberger et al., 2017; Alexander et al., 2020; Alexander et al., 2021), and for stratifying treatments (Prakash et al., 2021).

GNN studies, to date, have not made use of tau-specific positron emission tomography (PET) neuroimaging data, one of the key lesions in AD (Wong-Lin et al., 2020). Studies have suggested that tau PET brain images readily matched the distribution of tau deposits reported from histopathological studies, brain atrophy, hypometabolism and overall severity of AD (Saint-Aubert et al., 2017), and

might even be better than amyloid PET and MRI in AD prognosis (e.g. Ossenkoppele et al. (2021)). In terms of the studies using clustering, none have used the identified clusters for GNN-based AD classification.

Moreover, given that clinician diagnosis of AD can be subjective and not always consistent, there has been no utilization of more objective data-driven labelling of classes using unsupervised ML. Particularly, from an unsupervised cluster-based viewpoint, although previous studies demonstrated benefits in identifying sub-groups of patients, it is unclear, in the presence of tau PET neuroimaging data: (i) whether there are clear-cut AD clusters and if not, whether it is plausible to re-label AD class(es) for subsequent employment of GNN; and (ii) whether the GNN based on the re-labelled classes can perform sufficiently well as compared to GNN trained on clinician diagnostic labels.

In this study, we first apply nonlinear dimensional reduction on a heterogeneous dataset, which includes tau PET neuroimaging data co-registered with MRI. This is followed by data clustering, and then investigate the feature characteristics of the individual clusters. Next, we re-label cases based on cluster information, validated by tau PET data, to form new classes of AD and non-AD classes for GNN's AD classification. Finally, the GNN's performance using the re-labelled data is compared with that using clinician diagnosis.

2 MATERIALS AND METHODS

2.1 Data Description

The dataset used in this study was obtained from the open Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu), particularly the ADNIMERGE-3 open repository. Only ADNI participants who had undergone MRI and tau PET scans (for detecting tau deposition) were extracted from the data, and the PET and MRI PET data were merged with sociodemographic, medical/family history and neuropsychological features as measured at study baseline. The final dataset comprised of 224 features: 7 sociodemographic and medical history features, 40 cognitive and functional assessments' (CFAs) scores, and 177 neuroimaging features (from combined MRI and tau PET imaging data; see below). Clinician diagnosis of participants, considered as a class label for training GNN model (see below) consisted of control normal (CN), Alzheimer's Disease (AD), and mild cognitive impairment (MCI – which includes prodromal stage of AD).

The sociodemographic and medical/family history features were selected from each participant's medical and sociodemographic profile. These features were age, gender and years of education, maternal and paternal family history of AD, number of copies of the APoE ε4 alleles (abbreviated as APoE4). Reprocessed using 1-hot encoding, the APoE4 feature can take a value of 0, 1 or 2, representing the number of copies of the APoE4 allele. The CFA scores were collated from Alzheimer's Disease Assessment Scale (ADAS), Cognitive Battery Assessment, Clinical Dementia Rating (CDR), Mini-Mental State Exam (MMSE), Modified Hachinski Ischemia Scale, Neuropsychological Battery Test, logical memory immediate recall test (LMIT), logical memory delayed recall test (LMDT), the Neuropsychological Inventory (NPI) and the Geriatric Depression Scale (GDS). As well as total scores for all these assessments, individual question scores from ADAS and individual subscales from NPI were included in the dataset. Other CFAs and individual CFA subscales from ADNI were not included due to the large amounts of missing data for the set of participants who had undergone PET-MRI scans. For tau PET neuroimaging data, the [^{18}F]AV-1451 tracer for detecting tau deposition was used (Ossenkoppele et al., 2021), which has been recently approved by the U.S. Food and Drug Administration (FDA) (Jie et al., 2021).

In total, a dataset that comprised 559 samples representing individual participants was analysed, after data merging and pre-processing (see below). **Table I** shows a statistical summary of some sociodemographic and clinical data, with comparable average ages and years of education across the 3 diagnostic classes of CN, MCI and AD.

TABLE 1. Summary of the sociodemographic and clinical data.

Clinician diagnosis	Total	Age (Years)	Gender (M/F)	Education (Years)
CN	363	74.4 ± 7.3	150 / 213	16.84 ± 2.29
MCI	137	77.4 ± 8.15	84 / 53	16.13 ± 2.77
AD	59	77.8 ± 9.42	32 / 27	15.89 ± 2.57

Left column: Clinically diagnosed control normal (CN), mild cognitive impairment (MCI – which includes prodromal stage of AD) and Alzheimer's Disease (AD).

2.2 Data Preparation

Prior to data pre-processing, sociodemographic, medical/family history, CFAs and neuroimaging data were combined using each participant's unique identifier (ID), which ensured that every neuroimaging datapoint was merged with its respective participant profile. Inevitably, some participants were not present for a portion of the assessments without MRI and PET brain scans, which resulted in missing values. Although the dataset was well accounted for, outliers persisted as a direct result of those missing values. Relatively simple and sound imputation techniques were adopted in favour of more technical approaches as they tend to provide competitive performance with the absence of the computational and technical complexity (McCombe et al., 2022). Specifically, rows that contained sporadic missing values were imputed by means of single imputation technique (Ambler et al., 2007; Zhang, 2016), where: (i) categorical values were imputed using the mode of its respective feature; and (ii) continuous values were imputed using the mean of its respective feature.

However, participants missing a substantial portion of their cognitive assessment and medical/family history were imputed using a similarity matrix (Moeur and Stage, 1995; Boriah et al., 2008). It is important to note that these missing values pertained only to the sociodemographic portion of the feature set, affecting 6% of the entire dataset in total. This similarity matrix was calculated by taking the target participant with missing values and computing the distance between the target point and each remaining row. As the dataset was largely continuous, the Minkowski distance, in particular, the Frobenius norm (L_2 norm), was employed to calculate the distance between each instance and the target row (distance being a function of the coordinate vectors for each participant) (Moeur and Stage, 1995; Boriah et al., 2008). Approximately 10 members that showed close resemblance to the target row were provided by the resulting matrix, ultimately yielding a participant of greatest similarity to use as a reference for imputation.

Once all missing data imputation had been completed, the dataset was normalised. Negative columns were initially isolated and normalised by increasing all values in the column by the absolute of the minimum value in that column, as described by:

$$x' = x + |min(x)| \ . \tag{1}$$

for some variable or feature x. Lastly, all features were scaled using min-max normalisation:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \ . \tag{2}$$

2.3 Tau PET and MRI Data Pre-processing

The pre-processing steps for the PET and MRI data are outlined as follows. Each PET scan was co-registered with its associated MRI scan before normalising to a predetermined AD template, whereby the resulting images were corrected for partial volume estimation (PVE) using the SPM toolbox PETPVE12 (Müller-Gärtner et al., 1992; Gonzalez-Escamilla et al., 2017). This reference serves as a template for dictating measures from the aforementioned MRI scans. The values were sampled from brain regions defined by the Desikan-Killiany atlas (Desikan et al., 2006) where the resulting labels were formatted using this custom template. Ultimately, this process yielded 177 neuroimaging features per participant based on their combined PET-MRI data. **Figure 1** shows the distributions of tau levels for two sample brain regions, namely the (white-matter) right entorhinal cortex and the left amygdala for clinically diagnosed classes (CN and MCI combined, and AD), and all participants.

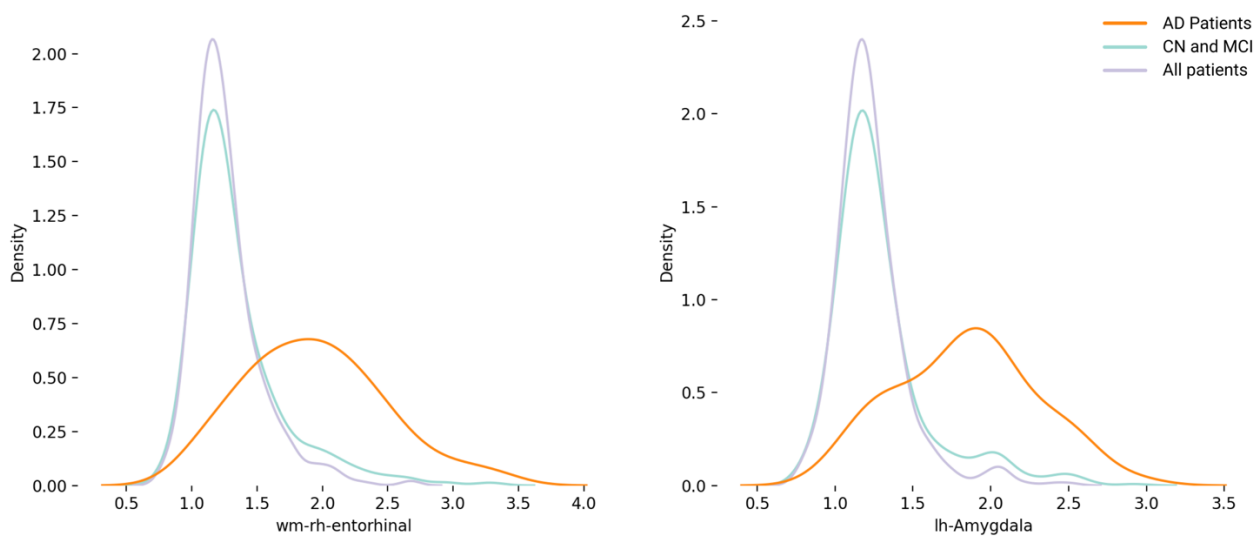

FIGURE 1. Distributions of partial volume estimation (PVE) for two sample brain regions (parcellation): right (rh) entorhinal cortex and white matter (wm), left (lh) amygdala regions according to participants' class (AD) or classes combined (CN, MCI and All). These are 2 out of a total of 177 brain PET-MRI features analysed.

2.4 Unsupervised Learning, Feature Selection and Class Re-labelling

After data normalisation (see Section 2.2), the dataset of 224 features and 559 samples was subjected to dimension reduction using the uniform manifold approximation and project (UMAP) for dimension reduction and data visualisation in lower dimensional space (McInnes et al., 2018). Unsupervised manifold learning allows for efficient embedding of non-linear data points while maintaining the relative distance or local connectivity of those points with respect to one another. By using an exponential probability distribution, as opposed to Euclidean distances, UMAP offers flexibility in that any distance function can be applied.

For this study, a UMAP cluster was implemented with a large nearest neighbour parameter (30 neighbours) (5) to avoid focusing on very local structures — a minimal distance value (0) was set to improve cluster density. 5 dimensions of the UMAP space was selected, based on visual inspection of the compactness of AD diagnosis cases distributed along these dimensions (see Results). Then, 3-dimensional UMAP plots were presented for data visualisation and clustering purposes. Each data point in each UMAP cluster was first assigned its original label (CN, MCI or AD) based on clinician diagnosis before having some of the data points re-labelled (see below).

Next, we used the unsupervised learning, *k*-means clustering (Lloyd, 1957; Forgy, 1965; Lloyd, 1982), to identify discrete clusters within the UMAP space. *k*-means clustering had also been applied successfully on AD datasets in previous studies (Martí-Juan et al., 2019; Alexander et al., 2020; Alexander et al., 2021), including the ADNI dataset. Characteristics of the clusters were extracted by using the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm (Cohen, 1995) as implemented in the JRip package in RWeka (Hornik et al., 2020) to identify the feature characteristics associated with membership of a particular cluster. Additionally, feature selection by the information gain algorithm (Larose and Larose, 2014) was implemented with the FSelector package in R (Romanski et al., 2021) to identify the top 10 features most associated with membership of a particular cluster for the originally labelled data and later, the re-labelled data. Once the key feature characteristics of each cluster were identified, the clusters were each given an appropriate unique name.

One of the clusters based on UMAP and *k*-means processing was subsequently identified to uniquely consist of a majority of clinically diagnosed AD cases; hence named the "AD" cluster. However, there were a few clinically diagnosed CN and MCI cases in this cluster, and some clinically diagnosed AD cases outside of the cluster. We re-labelled the CN and MCI cases in Cluster AD as AD cases while re-labelling AD cases in the other clusters as MCI cases. These re-labelled cases were validated post-hoc based on their tau PET levels, which were found to be similar, while being intermediate between the non-re-labelled AD and non-AD cases. The re-labelled data was used for training the GNN model, which will be compared to GNN trained using the original clinician diagnosis labels (see Section 2.5).

2.5 Graph Neural Network (GNN)

The robust meta-learning-based auto-metric graph neural network (AMGNN) classifier developed by Song et al., (2021) was used to classify the data using separately the original clinician diagnostic class labels and the new class labels. Song et al. (2021) used a few features of the data to create the graph for the neural network classifier, with the rest of the data features processed in the context of the graph relationships. In this work, the gender, age, education and family history data were used to build the graph for the classifier. The node classification of this small graph was done by randomly selecting samples from the training dataset as a meta-task to train the AMGNN. Based on several meta-task training runs, the AMGNN model can then be used to classify unknown label nodes in a new graph (Song et al., 2021). Overall, the network consisted of 2 GNN layers, alongside a total of 4 CNN layers.

The graph structure initialisation was the same as in Song et al. (2021) except that the data features and their number of features were different. Specifically, a fully connected graph was initialized to a given dataset such that $G = \{V, E\}$ where V represents the vertex set (participants), and E represents a set of pairs of vertices (graph edges). This was accompanied by the weights of the edge matrix, being assigned a value of *1*. Each node v_i was represented by the following set of features (Song et al., 2021):

$$v_i = [y_i, \eta_i, c_i, p_i] \tag{3}$$

where y_i defines the numeric classification of each label, η_i represents the risk factors (e.g. age, gender, years of education), while c_i represents the CFA scores. Lastly, p_i represents the tau PET and MRI feature set. The relationship or edge that exists between the two nodes v_i and v_j was denoted by e_{ij}. The construction of an AMGNN layer, loss function definition, and training strategy were the same as in Song et al. (2021). For further details, please refer to Song et al. (2021).

2.6 GNN Training and Testing

Using 559 resulting nodes, a small graph was constructed using batches of 64 randomly selected samples. This was repeated for both the supervised clinician's labels and cluster-based class label revisions of the dataset. The total number of training iterations was 250. Small batches were used to reduce both graph and therefore the computational complexity; this allowed for many iterations to be executed quickly. The dataset was split into training and testing datasets, at 80% and 20%, respectively. The model was trained using data batches of 64 rows before updating the loss parameter. Model performance was tested using one-shot learning using a single training sample in each batch (Garcia and Bruna, 2018). This was repeated for both datasets; one with original clinical diagnosis labels and the other based on cluster-based re-labelling. Model comparison and 3-class classification accuracy performance were visualized throughout training and testing process using TensorBoard, a visualization framework. Finally, the experiment was repeated 10 times to check for robustness and variability in performance. The rate of convergence during training sessions was also assessed.

3 RESULTS

3.1 Distinctive Data Cluster Characteristics in Low Dimensions

Using the dimensional reduction UMAP method (see Section 2.4), the initial dataset of 224 features were projected to 5 UMAP dimensions (**Figure 2A**). In the figure, the classes CN (dark purple), MCI (light purple) and AD (yellow) labelled by clinician diagnosis are denoted in different colours. A 5-dimensional UMAP space was selected, as visual inspection of the AD diagnosis information as distributed along the 5 dimensions demonstrated that the AD cases were clustered tightly together along some of these dimensions (**Figure 2A**, yellow points). A simpler 3-dimensional UMAP projection is illustrated in **Figure 2B**. Using *k*-means clustering, it can be observed that there were 5 distinct clusters (denoted by different colours). If we indicated the data points with clinical diagnosis (CN, MCI or AD), we found that the distribution of clinical diagnosis did not conform well to the 5 discrete clusters (**Table 2**).

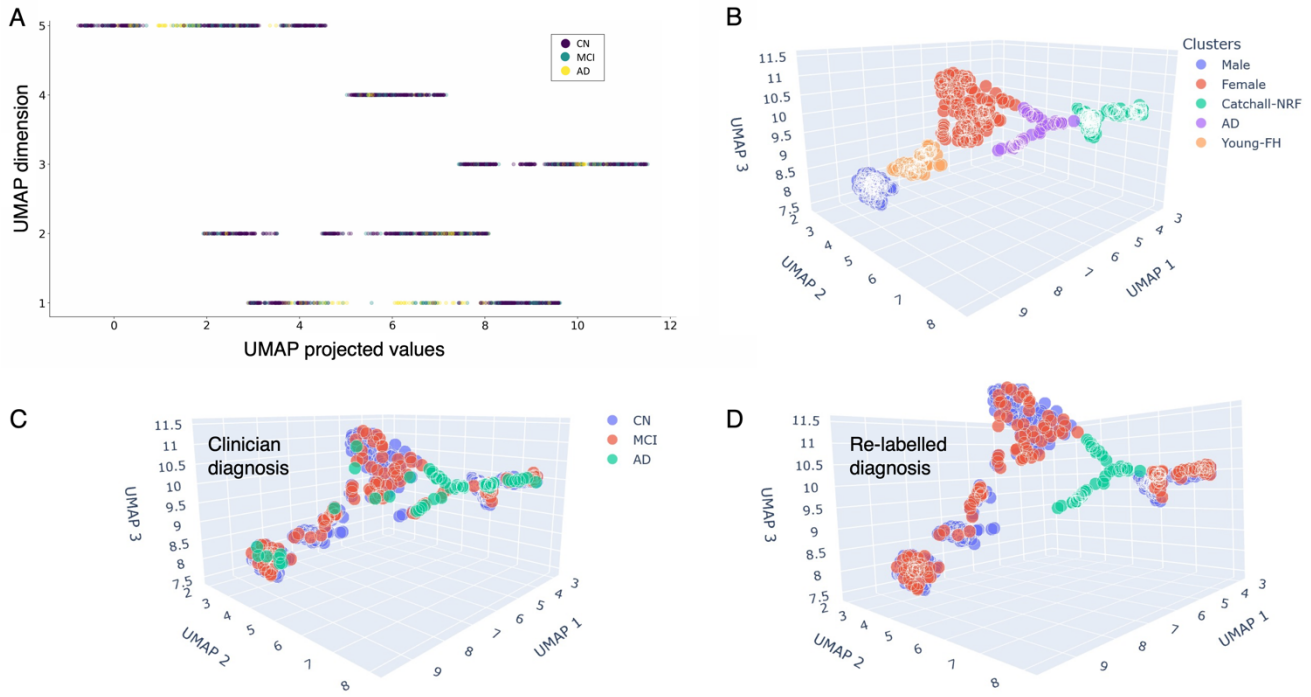

FIGURE 2. Data clusters in UMAP spaces. **(A)** Original data projected onto UMAP's 5 dimensions with distribution of projected values. Colour labelling of classes (clinician diagnosis): CN (dark purple), MCI (light purple) and AD (yellow). **(B)** Data clusters in 3-dimensional UMAP space. Different colours to indicate the distinct clusters determined by *k*-means clustering, association rule mining, information gain and tau PET imaging features. Cluster names were based on the unique feature characteristics of each cluster. **(C-D)** 3-dimensional UMAP with original clinician diagnostic labels **(C)** and with cluster-based class re-labelling **(D)**. Colour labelling for **(D)** are as in **(C)**.

Next, we sought to understand the characteristics of each cluster. Here, we used an association rule-based classifier generated by the RIPPER algorithm (Section 2.4) to understand the association of the features with each cluster. **Table 2** shows a summary of the sociodemographics and distribution of clinician diagnosis in each of the 5 clusters. In terms of years of education, all the clusters had similar average values around 16-17 years. The 2 most distinctive clusters were the gender specific clusters in which only male or female cases exist. This was consistent with previous unsupervised learning studies using different data (Gamberger et al., 2016; Alexander et al., 2021). Hence, we named these two clusters as Male and Female (**Table 2**). Interestingly, the Male cluster had the oldest average age among all clusters and had a higher proportion of MCI cases than the Female cluster. In comparison, there was a cluster with the youngest average age (71 years old) with about equal proportion of male and female participants, and which also had the highest proportion of CN cases. For now, we named this cluster the Young cluster. Another distinctive cluster is one with the largest proportion of AD cases (52%), despite not having the oldest average age. We named this cluster as the AD cluster even though MCI cases constituted 36%. The final clusters consisted of almost equally mixed gender and with a substantially high number of CN cases. For now, we named this the Catchall cluster.

TABLE 2. Summary of the 5 clusters' sociodemographics and clinician diagnosis distribution.

Cluster Label	**No. of Cases**	**Age (Years)**	**Gender: M/F (%, in whole number)**	**Education (Years)**	**Clinician Diagnosis: CN/MCI/AD (%, in whole number)**
AD	50	75.1 ± 9.8	30/20 (60/40)	16.7 ± 2.5	6/**18**/**26** (12/**36**/**52**)
Catchall	166	74.6 ± 7.2	70/96 (42/58)	16.3 ± 2.59	**114**/32/20 (**69**/**19**/12)
Female	131	75.9 ± 8.2	0/**131** (0/**100**)	16.0 ± 2.4	**96**/**28**/7 (**73**/**21**/5)
Male	131	**78.8** ± 6.9	**131**/0 (**100**/0)	17.1 ± 2.5	**82**/**44**/5 (**63**/**34**/4)
Young-FH	81	**71.3** ± 6.4	35/46 (43/57)	17.1 ± 2.1	**65**/**15**/1 (**80**/**19**/1)

By successively applying the rules of the association rule-based classifiers (Section 2.4), it is possible to discern some other features that distinguish each cluster. **Table 3** summarises the results for the 5 clusters. In total, there were 7 association rules found. First, the algorithm revealed that if there were 2 copies of the APoE ε4 allele, a known risk genetic factor (e.g. Angelopoulou et al., 2021), then 90% of the cases would belong to Cluster AD. We next apply the classifier on the remaining cases, and observed that if there was high tau level in the left fusiform cortex with no APoE ε4 allele, then the algorithm predicted about 94% of the data to again be in Cluster AD. The fusiform brain region is known to be involved in cognitive function, particularly facial recognition, and AD patients have issues with recognising faces (e.g. Huang et al., 2020, 2021); hence this was a consistent funding.

Interestingly, demographics and medical/family history also played a role in cluster membership. After applying the first two rules, if the remaining participants' mother had AD but no APoE ε4 allele, then there was ~96% chance of being in Cluster Young; we renamed this cluster as Young-FH (FH: family history). Applying the classifier again uncovered that if the participants were female with no APoE4 ε4 allele, then about 99% of the time could be found in Cluster Female.

For the remaining cases, when there was low level of tau in the left-entorhinal white-matter with no APoE4 ε4 allele, there was 100% chance of being in Cluster Male. It should be noted that the entorhinal brain region is associated with early-stage AD (Braak and Braak, 1991, 1996, 1997a, 1997b). Finally, for the rest of the data, about 98% belonged to Cluster Catchall.

TABLE 3. Summary of the association rules identified for distinguishing the data clusters over the algorithmic steps.

Steps	Association Rules	Inside/Outside Cluster (%, in whole number)
1	(APoE4 = 1) ⇒ Cluster AD	36/4 (90/10)
2	(ctx.lh.fusiform ≥ 1.777) and (APoE4 = 0) ⇒ Cluster AD	15/1 (94/6)
3	(Mother with AD) and (APoE4 = 0) ⇒ Cluster Young-FH	79/3 (96/4)
4	(Female) and (APoE4 = 0) ⇒ Cluster Female	129/130 (99/1)
5	(ctx.lh.entorhinal ≤ 1.977) and (APoE4 = 0) ⇒ Cluster Male	131/0 (100/0)
6	(all remaining cases) ⇒ Cluster Catchall	169/4 (98/2)

We next further analysed the clusters' characteristics using a more extensive feature selection method – information gain of the data features with respect to cluster labels (Section 2.4). In particular, we hoped that this method would shed light on the Catchall cluster. The top ranked features found by information gain that could distinguish the clusters were (ranked from the most important feature): APoE4, gender, (history of) mother with AD, ctx.rh.inferiorparietal, ctx.lh.middletemporal, lh.Amygdala, ctx.lh.inferiorparietal, ctx.lh.lateraloccipital, wm.lh.entorhinal, and wm.lh.inferiortemporal. Consistent with the association rule mining results, APoE4, gender and (history) of mother with AD were identified as the top 3 features. Moreover, the list was dominated by various tau PET-MRI imaging features (7 out of 10 features). Hence, we analysed the clusters using the ctx.lh.inferiorparietal (left inferior parietal cortex) and lh.Amygdala (left amygdala) features, based on their suggested links to early stages of MCI and AD (Greene et al., 2010; Coupé et al., 2019).

Table 4 summarises the statistics of participants of the values for the ctx.lh.inferiorparietal and lh.Amygdala features across the 5 clusters. We can see that Cluster AD consistently had the largest values for all the statistics regardless for both brain regions (**Table 2**, bold text), as was expected. Interestingly, Cluster Catchall had the second largest values for most of the statistics (**Table 2**, bold text). These could be due to the 12% of clinically diagnosed AD cases (second highest cases among all the clusters), albeit 69% CN and 19% MCI cases (**Table 2**). Hence, this cluster might consist of early MCI or AD stage, with or without formal diagnosis by clinicians – i.e. with potential neurological risk factor (NRF). We therefore renamed this cluster as Cluster Catchall-NRF.

TABLE 4. Summary of the statistics of participants' tau PET-MRI neuroimaging features, ctx.inferiorparietal and lh.Amygdala, with respect to the 5 clusters. Values shown up to 3 decimal places. Bold: Prominently high values across each column and each brain region.

Cluster	Mean	Standard Deviation	Minimum	1st Quartile	Median	3rd Quartile	Maximum
ctx.lh.inferiorparietal							
AD	**2.363**	**1.277**	**1.211**	**1.541**	**1.751**	**2.759**	**6.193**
Catchall-NRF	**1.506**	**0.362**	0.949	1.317	**1.422**	**1.573**	**3.139**
Female	1.438	0.212	1.070	1.327	1.415	1.518	3.014
Male	1.319	0.164	0.970	1.220	1.289	1.410	1.992
Young-FH	1.395	0.160	1.028	1.299	1.377	1.467	2.137
lh.Amygdala							
AD	**1.831**	**0.439**	**0.969**	**1.522**	**1.812**	**2.080**	**2.918**
Catchall-NRF	**1.336**	**0.342**	0.816	**1.117**	**1.245**	**1.424**	**2.590**
Female	1.185	0.185	0.758	1.104	1.167	1.271	2.108
Male	1.198	0.160	0.826	1.090	1.204	1.286	1.731
Young-FH	1.196	0.165	0.965	1.078	1.174	1.263	1.746

3.2 Re-labelling of Classes

When clinician diagnosis (CN, MCI or AD) was indicated in these clusters (**Figure 2C**), we observed overlaps between the UMAP clusters and the clinician diagnostic labelled classes, with substantial mixing between the classes in some of the clusters. As mentioned earlier, one of the clusters (Cluster AD) particularly had substantially purer (AD) cases. Hence, we hypothesised that we could re-label the CN and MCI cases in Cluster AD as AD cases (CN/MCI-to-AD), while re-labelling AD cases in

the other clusters as MCI cases (AD-to-MCI). The re-labelled data points are visually shown in **Figure 2D** (compared to **Figure 2C**).

To validate this, we investigated post-hoc the tau PET-MRI features of these re-labelled cases to check whether they could together be considered as a form intermediate AD stage with difficulty of diagnosis. Before doing that, we again used information gain to identify the top ranked features with respect to cluster labels of the re-labelled data, and we found the following features (ranked from the most to least important): CFA harmonised composite memory (PHC_MEM), ctx.lh.entorhinal, another composite memory scale ADNI_MEM, wm.lh.entorhinal, lh.Amygdala, wm.lh.inferiortemporal, rh.Amygdala, wm.rh.inferiortemporal, ctx.lh.inferiortemporal and wm.lh.fusiform. Compared to the original dataset (see Section 3.1. above), only lh.Amygdala and wm.lh.inferiortemporal features were common.

Next, based on the selected features we made use of PHC_MEM, lh.Amygdala and and wm.lh.entorhinal feature to perform a post-hoc check for the re-labelled cases, given the known link of the amygdala to early-stage AD (Dickerson et al., 2001; Killiany et al., 2002; Squire et al., 2004; Coupé et al., 2019). **Table 5** shows the mean and standard deviation for these 2 features for the 4 cases: (i) relabelled CN/MCI-to-AD in Cluster AD; (ii) AD-to-MCI outside Cluster AD; (iii) remained AD cases in Cluster AD; and (iv) remained non-AD cases outside Cluster AD. We can see that for the AD-to-MCI (outside Cluster AD) and CN/MCI-to-AD (in Cluster AD) re-labelled cases, the values for both the lh.Amygdala and wm.lh.entorhinal features' values were relatively similar, while lying intermediate between the values of the remained AD cases and remained non-AD cases. This neuroimaging features supported the re-labelling hypothesis regarding the clinician diagnostic uncertainty. With regards to the CFA, PHC_MEM, the pattern was not as clean. Specifically, the values for the AD-to-MCI (outside Cluster AD) and CN/MCI-to-AD (in Cluster AD) re-labelled cases were not too close to each other, albeit both have mean values to lie between the values for the remained AD and remained non-AD cases. In fact, the AD-to-MCI cases seemed to have values leaning closer towards that of the remained AD cases.

TABLE 5. Summary of statistics of participants' neuroimaging features lh.Amygdala and wm.lh.entorhinal with respect to the relabelled CN/MCI-to-AD in Cluster AD, and AD-to-MCI outside Cluster AD (bold text), and compared them to those of remained AD cases in Cluster AD and remained non-AD cases outside Cluster AD; hence 4 cases to consider. Values shown up to 3 decimal places.

Types of Cases	**No. of Cases**	**PHC_MEM (Mean, Standard Devation)**	**lh.Amygdala (Mean, Standard Deviation)**	**wm.lh.entorhinal (Mean, Standard Deviation**
AD-to-MCI (outside Cluster AD)	33	**-0.864±0.585**	**1.738±0.432**	**1.815±0.508**
Remained non-AD cases (outside Cluster AD)	476	0.712±0.603	1.205±0.188	1.233±0.221
CN/MCI-to-AD (in Cluster AD)	24	**0.079±0.571**	**1.677±0.399**	**1.708±0.348**
Remained AD cases (in Cluster AD)	26	-1.092±0.521	1.974±0.433	2.101±0.477

3.3 GNN based Training and Classification of CN, MCI and AD

With the re-labelled data, we can now train the GNN model by Song et al. (2021) (Sections 2.5-2.6) using both the originally labelled data based on clinician diagnosis and the re-labelled data (in Section 3.2). We then check for the trends of convergence for the GNN models across all 30 experimental repetitions. We found that the trend of convergence for the GNN model using the re-labelled data to be faster than when using the originally labelled data. The convergence trends for the first 10 repetitions of each experiment are shown in **Figure 3**. Specifically, the GNN model using the re-labelled data reached a 3-class classification accuracy of ~90% by the ~80th iteration while the GNN model using the originally labelled data took ~120 iterations to reach the same accuracy level. Over the 30 repetitions, the GNN model using the re-labelled data achieved an average accuracy of 93.20±0.03%, which should be compared to that using the originally labelled data of 90.06±0.04% accuracy. A t-test comparing accuracy over the 30 repetitions yielded a p-value of 0.011. Hence, not only was the cluster-labelled based model's average accuracy significantly higher, but its performance was less variable. This could be related to the original classification of the labels, with perhaps the labels following a less predictable pattern.

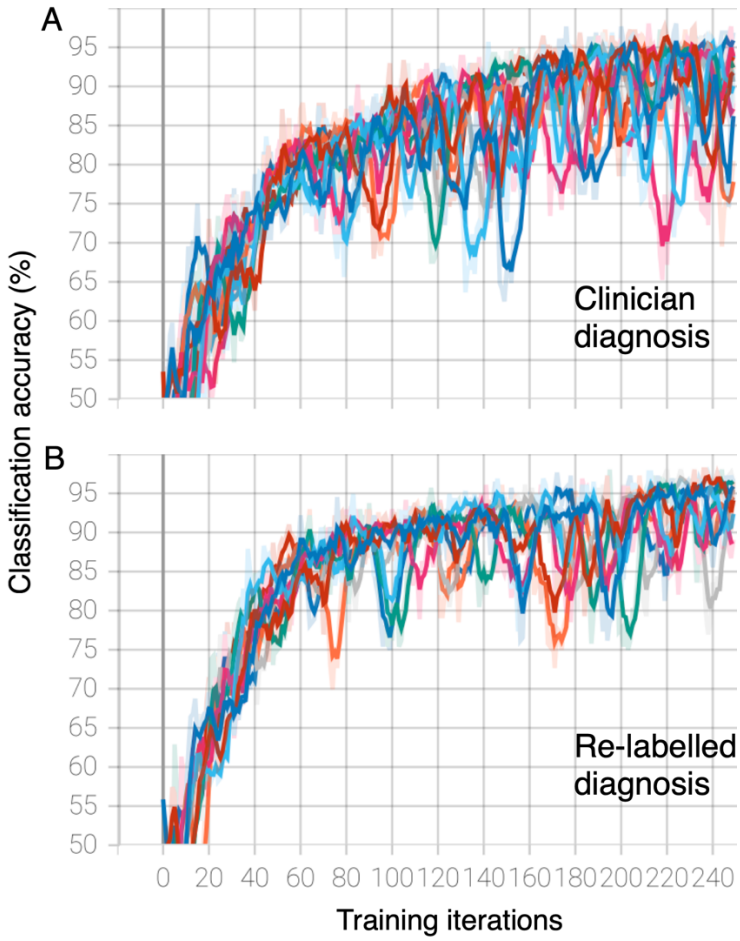

FIGURE 3. GNN (3-class) classification accuracy across training iterations or epochs with 10 repetitions. **(A-B)** Convergence for GNN model using originally labelled **(A)** and re-labelled **(B)** data.

4 DISCUSSION

This study has successfully demonstrated the application of a GNN model on heterogenous dementia data types using both clinician diagnostic labels and cluster-based labels. The heterogeneous data consisted of combined tau PET neuroimaging and MRI data, together with sociodemographic data, medical/family history and cognitive and functional assessments. The GNN model using the re-labelled data converged faster with higher stability during training, while achieving significantly higher (3-class) classification accuracy than that using the original clinician diagnostic labels.

GNNs have been introduced into AD classification studies in a variety of ways using different types of data, especially MRI and PET brain data (Parisot et al., 2018; Kazi et al., 2019a; Kazi et al., 2019b; Wen et al., 2020; Vivar et al., 2021; Zhu and Razavian, 2021; Wang et al., 2021; Yang et al., 2021; Bessadok et al., 2021). More recent advancements of GNNs on AD have been proposed to provide more flexibility (Zhu and Razavian, 2021; Song et al., 2021). However, these studies have not considered detailed tau PET neuroimaging data, a limitation given the closer alignment of tau PET with AD stages (Saint-Aubert et al., 2017; Ossenkoppele et al., 2021), than e.g. amyloid PET, and the recent approval of its use by the U.S. FDA (Jie et al., 2021). Importantly, these studies, together with the literature on unsupervised learning approaches applied to AD (Gamberger et al., 2016; Gamberger et al., 2017; Martí-Juan et al., 2019; Ferreira et al., 2019; Alashwal et al., 2019; Wang et al., 2021;

Katabathula et al., 2021; Prakash et al., 2021, Alexander et al., 2020; Alexander et al., 2021; Prakash et al., 2021), have not used cluster-based class re-labelling for GNN's classification of AD. Here, we have made use of the robust auto-metric GNN model by Song et al. (2021). Unlike Song et al. (2021) who used a more limited (TADPOLE) dataset, this work made use of the more extensive ADNIMERGE-3 dataset with individuals that had tau PET brain scans. Moreover, our work included more cognitive and functional assessments. Importantly, we made use of UMAP-cluster based re-labelling to train the GNN model.

Prior to applying GNN for AD classification, we made use of nonlinear dimensional reduction UMAP (**Figure 2**) and projected the data into a lower 3-dimensional UMAP (**Figures 2B-D**) space for data visualisation for deeper insights and for guidance in the re-labelling (**Figure 2D**). Five discrete data clusters were identified using *k*-means clustering (**Figure 2B**). Applying association rule mining successively to the data, we found key feature characteristics (**Tables 2** and **3**) uniquely defining four of these five clusters, namely, a majority AD cluster, a fully male cluster, a fully female cluster, a cluster of younger participants with parental history of AD ("Young-FH" cluster), and a relatively unknown "Catchall" cluster. It is interesting to speculate that the Young-FH could be due to relatively younger participants who were concerned about their own health given that their parents had a history of AD. The 2 gender specific clusters identified seemed to be in line with previous studies (Gamberger et al., 2016; Prakash et al., 2021, Alexander et al., 2020; Alexander et al., 2021). Cluster AD membership was associated with having 2 copies of APoE ε4 allele or had high tau level in the left fusiform cortex, consistent with genetic risk factor (e.g. Angelopoulou et al., 2021) and late-stage AD brain changes (e.g. Huang et al. 2020, 2021), respectively. The substantially higher number of AD cases in this cluster provides hope that the unsupervised learning's clustering approach might reveal something interesting about the disease while helping to guide the subsequent re-labelling of diagnostic classes.

Further inspection using information gain feature selection identified the top 10 most important features identified within the 5 clusters. Other than the top 3 features (APoE4, gender and parental history of AD), the remainder were tau PET-MRI imaging features. Based on their suggested links to early stages of MCI and AD (Greene et al., 2010; Coupé et al., 2019), we selected 2 of these tau PET-MRI features, the ctx.lh.inferiorparietal (left inferior parietal cortex) and lh.Amygdala (left amygdala) features, to provide further insights into the clusters. These regions were chosen due to their suggested links to early-stage MCI/AD (Greene et al., 2010; Coupé et al., 2019). It turned out that the previously named Catchall cluster had levels of tau between those of the AD cluster and the 3 other clusters (**Table 4**). This might indicate that this cluster could constitute cases with underlying neurological risk factors (NRF), which we later renamed as the Catchall-NRF cluster.

Next, we re-labelled the diagnostic classes based around the AD cluster such that AD cases occurred only within this cluster while those AD cases outside of it were re-labelled as a prodromal stage of AD. Specifically, we re-labelled the CN and MCI cases in Cluster AD as AD cases (CN/MCI-to-AD), while re-labelling AD cases in the other clusters as MCI cases (AD-to-MCI). This process was supported by post-hoc analysis of 2 tau PET-MRI features (lh.Amygdala and wm.lh.entorhinal) and 1 CFA feature (PHC_MEM), obtained from information gain feature selection of the re-labelled data, revealing similar brain tau PET levels for the CN/MCI-to-AD and AD-to-MCI cases, and lying between those of the remained AD and non-AD cases. These 2 brain regions were based on their suggested links to early-stage AD (Dickerson et al., 2001; Killiany et al., 2002; Squire et al., 2004; Coupé et al., 2019). These re-labelled data led to more accurate and stable GNN model for detecting CN, MCI and AD groups. In comparison, although the PHC_MEM although follow similar trend, the latter was not as clear-cut. This showed that CFAs might not be as well related to the clusters. This was quite expected

as the overall features leading to the clusters were dominated by tau PET-MRI neuroimaging features, and that neuroimaging features were more objective than CFAs.

Taken together, we have successfully incorporated both unsupervised learning (UMAP, *k*-means clustering, association rule mining) and supervised learning (information gain feature selection) to provide insights into tau PET-MRI and other non-neuroimaging heterogeneous data, and subsequently using supervised learning (GNN) for detecting CN/MCI/AD. Importantly, combined UMAP and clustering highlighted new facets in terms of features that can be used for AD diagnosis – in an attempt to deconstruct the current standards for determining features actively contributing to AD. Ultimately, using clustering analysis provides more emphasis in exploring unknown AD contributors as opposed to accepting the current status-quo. It should, however, be noted that the re-labelling of the data required not just algorithmic approaches, but also information from clinical diagnosis of AD. Indeed, we were fortunate to identify a data cluster to comprise mainly of AD cases, even though the MCI cases were rather more distributed. Future work should investigate situations where diagnostic cases were more distributed, and to distinguish the subtypes of MCI and AD cases and their progression (Gamberger et al., 2017; Ferreira et al., 2017, 2019, 2020; Alexander et al., 2020; Alexander et al., 2021; Mitelpunkt et al., 2020; Vogel et al., 2021).

The granularity of the data used to train the GNN model could be expanded to derive new patterns governing AD development. For instance, this could involve inclusion of blood- or cerebrospinal-based biomarkers, or with a larger number of PET scan participants in the ADNI data it could include the breaking down of additional cognitive and functional assessments into their smaller components, as specific questions in a neuropsychological questionnaire (McCombe et al., 2021). In fact, it would be worth investigating using larger dataset with PET-MRI features to validate our clustering and AD-classification findings. As the ADNI study continues it may also become possible to look at the PET-MRI data longitudinally, for example, to ascertain whether members of the NRF cluster are at increased risk of cognitive decline. It would also be of great value to explore the decision-making processes of GNNs in more depth. For example, Wen et al. (2020) has provided extensive discussion in the ways of GNN modelling and feature interpretation – processes which could be adapted to the GNN modelling techniques utilised in this study to further improve the interpretability and accessibility of ML research in computer-aided diagnostic systems. Further opportunities exist in the use of MRI/PET data analysis within AD diagnosis. In our study, neuroimaging data was processed using PETPVE12 according to the Desikan-Killiany atlas (Desikan et al., 2006), however, the neuroimaging data could be processed in a variety of atlas standards which may permit the GNN model to provide different interpretations of the same underlying data, or perhaps even improve its diagnostic accuracy.

5 CONCLUSION

The aim of this study was to demonstrate the benefits of cluster-based labelling for graph neural network's multi-class classification of AD, using tau PET imaging together with other heterogeneous data. Original classes based on clinician diagnosis were used and cluster-based re-labelled classes were devised using latent features derived from UMAP and various analytical methods. The high diagnostic accuracy of the latter approach highlighted the potential of how the labelling of an individual having MCI or AD can be disrupted via less subjective or biased measures. This study reinforces the need for current AD diagnostic practices to continue to be re-evaluated by using more objective data-driven, e.g. unsupervised learning and clustering methods, to derive new patterns and sub-groups from existing datasets.

DATA AVAILABILITY STATEMENT

Python and R codes for clustering, data analysis and classification are available from the authors upon reasonable request. Requests to access the original datasets should be directed to ADNI (http://adni.loni.usc.edu/).

AUTHOR CONTRIBUTIONS

NM, JB and KW-L conceptualized this study. JMS-B processed the combined PET-MRI data. NM and JB developed the codes for dimensional reduction, and NM developed the codes for clustering and feature selection. JB developed the codes for graph neural network modelling, and NM and JB performed the analyses. NM, JB and KW-L interpreted the results and wrote the bulk of the document. DPF and PLM provided comments and insight into the analyses, and contributed to the editing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the European Union's INTERREG VA Programme, managed by the Special EU Programmes Body (SEUPB; Centre for Personalised Medicine, IVA 5036). The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB). The views and opinions expressed in this paper do not necessarily reflect those of the European Commission or the Special EU Programmes Body (SEUPB).

ACKNOWLEDGMENTS

We thank Alok Joshi for providing useful feedback on the work and an earlier version of the manuscript. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

REFERENCES

Alashwal, H., El Halaby, M., Crouse, J.J., Abdalla, A. and Moustafa, A. A. (2019). The application of unsupervised clustering methods to Alzheimer's disease. *Frontiers in Computational Neuroscience*, 13:31. doi: 10.3389/fncom.2019.00031

Alexander, N., Alexander, D., Barkhof, F. and Denaxas, S. (2020) Using unsupervised learning to identify clinical subtypes of Alzheimer's disease in electronic health records. *Studies in Health Technology and Informatics,* 270:499-503.

Alexander, N., Alexander, D.C., Barkhof, F. and Denaxas, S. (2021) Identifying and evaluating clinical subtypes of Alzheimer's disease in care electronic health records using unsupervised machine learning. *BMC Medical Informatics and Decision Making*, 21:343. doi: 10.1186/s12911-021-01693-6

Ambler, G., Omar, R. and Royston, P. (2007) A comparison of imputation techniques for handling missing predictor values in a risk model with a binary outcome. *Statistical Methods in Medical Research,* 16(3):277-298.

Angelopoulou, E., Paudel, Y.N., Papageorgiou, S.G., Piperi, C. (2021) APOE genotype and Alzheimer's disease: The influence of lifestyle and environmental factors. *ACS Chemical Neuroscience*, 12(15):2749-2764. doi: 10.1021/acschemneuro.1c00295.

Bessadok, A., M. Mahjoub and I. Rekik (2021) Graph neural networks in network neuroscience. arXiv preprint, pp. 1-27, 2021. arXiv: 2106.03535

Braak, H. and Braak, E. (1991) Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82:239-259.

Braak, H. and Braak, E. (1996) Development of Alzheimer-related neurofibrillary changes in the neocortex inversely recapitulates cortical myelogenesis. *Acta Neuropathologica*, 92:197-201.

Braak, H. and Braak, E. (1997a) Diagnostic criteria for neuropathologic assessment of Alzheimer's disease. *Neurobiological Aging*, 18:S85-S88.

Braak, H. and Braak, E. (1997b) Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiological Aging*, 18:351-357.

Burns, A., and Iliffe, S. (2009). Alzheimer's disease. *BMJ*, 338:b158. doi: 10.1136/bmj.b158

Cohen, W.W. (1995) Fast effective rule induction. *In: Machine Learning Proceedings* 1995, pp. 115-123. doi: 10.1016/B978-1-55860-377-6.50023-2

Coupé, P., Manjón, J.V., Lanuza, E. and Catheline, G. (2019) Lifespan changes of the human brain in Alzheimer's disease. *Scientific Reports*, 9:3998. doi: 10.1038/s41598-019-39809-8

Desikan, R., Ségonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B. and Albert, M. (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage,* 31(3):968-980.

Dickerson, B.C., Goncharova, I., Sullivan, M.P., Forchetti, C., Wilson, R.S., Bennett, D.A., Beckett, L.A. and deToledo-Morrell, L. (2001). MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease. *Neurobiology of Aging*, 22:747–754. doi: 10.1016/s0197-4580(01)00271-8

Ferreira, A.V.A., Bastos Filho, C.J.A. and Lins, A.J.C.C. (2019) An unsupervised analysis of an Alzheimer's disease patient population using subspace search and hierarchical density-based clustering. *In: 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2019, pp. 1-6, doi: 10.1109/LA-CCI47412.2019.9037028.

Ferri, C., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, K., Hasegawa, K., Hendrie, H., Huang, Y. Jorm, A., Mathers, C., Menezes, P., Rimmer, E., Scazufca, M., and Alzheimer's Disease International (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet,* 366, 2112-2117.

Ferreira, D., Verhagen, C., Hernández-Cabrera, J., Cavallin, L., Guo, C.-J., Ekman, U., Muehlboeck, J.-S., Simmons, A., Barroso, J., Wahlund, L.-O. and Westman, E. (2017) Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. *Scientific Reports*, 7:46263. doi: 10.1038/srep46263

Ferreira, D., Pereira, J.B., Giovanni, V. and Westman, E.; for the Alzheimer's Disease Neuroimaging Initiative (2019) Subtypes of Alzheimer's disease display distinct network abnormalities extending beyond their pattern of brain atrophy. *Frontiers in Neurology*, 10:524. doi: 10.3389/fneur.2019.00524

Ferreira, D., Nordberg, A. and Westman, E. (2020) Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology*, 94(10):436-448. doi: 10.1212/WNL.0000000000009058

Forgy, E.W. (1965) Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21(3):768–769.

Gamberger, D., Ženko, B., Mitelpunkt, A., Shachar, N. and Lavrač, N. (2016) Clusters of male and female Alzheimer's disease patients in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. *Brain Informatics,* **3:**169–179. doi: 10.1007/s40708-016-0035-5

Gamberger, D., Lavrač, N., Srivatsa, S., Tanzi, R.E. and Doraiswamy, P.M. (2017) Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease. *Scientific Reports*, 7:6763. doi: 10.1038/s41598-017-06624-y

Garcia, V. and Bruna, J. (2018) Few-shot learning with graph neural networks. *In: Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)*, pp. 1-13.

Gonzalez-Escamilla, G., Lange, C., Teipel, S., Buchert, R. and Grothe, M. (2017) Alzheimer's Disease Neuroimaging Initiative. PETPVE12: an SPM toolbox for Partial Volume Effects correction in brain PET - Application to amyloid imaging with AV45-PET," *NeuroImage,* 147: 669-677.

Gori, M., Monfardini, G. and Scarselli, F. (2005) A new model for learning in graph domains. *Proceedings. In: 2005 IEEE International Joint Conference on Neural Networks,* 2: 729-724.

Greene, S.J., Killiany, R.J. and Alzheimer's Disease Neuroimaging Initiative (2010) Subregions of the inferior parietal lobule are affected in the progression to Alzheimer's disease. *Neurobiology of aging*, 31(8):1304–1311.

Hornik, K., Buchta, C., Hothorn, T., Karatzoglou, A., Meyer, D. and Zeileis, A. (2020) R/Weka interface, *R package version* 0.4-43.

Huang, J., Beach, P., Bozoki, A. and Zhu, D.C. (2020) Alzheimer's Disease Progressively Alters the Face-Evoked Visual-Processing Network. *Journal of Alzheimer's Disease*, 77(3):1025-1042.

Huang, J., Beach, P., Bozoki, A., Zhu and D.C. (2021) Alzheimer's Disease Progressively Reduces Visual Functional Network Connectivity. *Journal of Alzheimer's Disorder Reports*, 5(1):549-562.

Jie, C.V.M.L., Treyer, V., Schibli and R., Mu, L. (2021) Tauvid™: The First FDA-Approved PET Tracer for Imaging Tau Pathology in Alzheimer's Disease. Pharmaceuticals (Basel), 14(2):110. doi: 10.3390/ph14020110

Katabathula, S., Wang, Q. and Xu, R. (2021) Predict Alzheimer's disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations.," *Alzheimer's Research & Therapy,* 13(1):1-9.

Kazi, A., Shekarforoush, S., Kortuem, K., Albarqouni, S. and Navab, N. (2019a) Self-attention equipped graph convolutions for disease prediction. *In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019),* pp. 1896-1899. doi: 10.1109/ISBI.2019.8759274

Kazi, A., Shekarforoush, S., Krishna, S., Burwinkel, H., Vivar, G., Wiestler, B., Kortüm, K., Ahmadi, S.-A., Albarqouni, S. and Navab, N. (2019b) Graph convolution based attention model for personalized disease prediction. *In: Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. MICCAI 2019. Lecture Notes in Computer Science, 11767. Springer, Cham*. doi: 10.1007/978-3-030-32251-9_14

Killiany, R.J., Hyman, B.T., Gomez-Isla, T., Moss, M.B., Kikinis, R., Jolesz, F., Tanzi, R., Jones, K. and Albert, M.S. (2002). MRI measures of entorhinal cortex vs hippocampus in preclinical AD. *Neurology*, 58(8):1188–1196. doi: 10.1212/wnl.58.8.1188

Larose, D.T. and Larose, C.D. (2014) *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, New Jersey, Wiley.

Lloyd, S.P. (1957) Least square quantization in PCM. *Bell Telephone Laboratories Paper*.

Lloyd, S.P. (1982) Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129-137.

Martí-Juan, G., Sanroma, G., Piella, G.; Alzheimer's Disease Neuroimaging Initiative and the Alzheimer's Disease Metabolomics Consortium (2019) Revealing heterogeneity of brain imaging phenotypes in Alzheimer's disease based on unsupervised clustering of blood marker profiles. PLoS One, 14(3):e0211121. doi: 10.1371/journal.pone.0211121

McCombe, N., Liu, S., Ding, X., Prasad, G., Bucholc, M., Finn, D.P., Todd, S., McClean, P.L. and Wong-Lin, K. (2022) Practical Strategies for Extreme Missing Data Imputation in Dementia Diagnosis. *IEEE Journal of Biomedical and Health Informatics,* 26(2):818-827. doi: 10.1109/JBHI.2021.3098511

McCombe, N., Ding, X., Prasad, G., Gillespie, P., Finn, D.P., Todd, S., McClean, P.L. and Wong-Lin, K. (2021) Accuracy-time optimisation with cost-sensitive feature selection for Alzheimer's disease diagnosis. medRxiv preprint, pp. 1-9, 2021. doi:10.1101/2021.07.01.21259858

Mitelpunkt, A., Galili, T., Kozlovski, T., Bregman, N., Shachar, N., Markus-Kalish, M. and Benjamini, Y. (2020) Novel Alzheimer's disease subtypes identified using a data and knowledge driven strategy. *Scientific Reports*, 10:1327. doi: 10.1038/s41598-020-57785-2

Moeur, M. and Stage, A. (1995) Most similar neighbor: An improved sampling inference procedure for natural resource planning. *Forest Science,* 41(2):337-359.

Müller-Gärtner, H., Links, J., Prince, J., Bryan, R., McVeigh, E., Leal, J., Davatzikos, C. and Frost, J. (1992) Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects," *Journal of Cerebral Blood Flow & Metabolism,* 12(4):571-583.

Mukherjee, S., Mez, J., Trittschuh, E., Saykin, A., Gibbons, L., Fardo, D., Wessels, M., Bauman, J., Moore, M., Choi, S. and Gross, A. (2020) Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Molecular Psychiatry*, 25(11):2492-2951.

Ossenkoppele, R., Smith, R., Mattsson-Carlgren, N., et al. (2021) Accuracy of tau positron emission tomography as a prognostic marker in preclinical and prodromal Alzheimer disease: A head-to-head

comparison against amyloid positron emission tomography and magnetic resonance imaging. *JAMA Neurology*, 78(8):961–971. doi:10.1001/jamaneurol.2021.1858

Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B. and Rueckert, D. (2018) Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease. *Medical Image Analysis,* 48:117-130.

Prakash, J., Wang, V., Quinn, R.E. III and Mitchell, C.S. (2021) Unsupervised machine learning to identify separable clinical Alzheimer's disease sub-populations. *Brain Sciences* 11, no. 8: 977. https://doi.org/10.3390/brainsci11080977

Romanski, R., Kotthoff, L. and Schratz, P. (2021) FSelector, *R package version* 0.33.

Saint-Aubert, L., Lemoine, L., Chiotis, K., Leuzy, A., Rodriguez-Vieitez, E. and Nordberg, A. (2017) Tau PET imaging: present and future directions. *Molecular Neurodegeneration*, 12:19. doi: 10.1186/s13024-017-0162-3

Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G. (2009) The graph neural network model. *IEEE Transactions on Neural Networks,* 20(1):61-80.

Squire, L.R., Stark, C.E.L. and Clark, R.E. (2004). The medial temporal lobe. *Annual Reviews of Neuroscience*, 27:279–306.

Song, X., Mao, M. and Qian, X. (2021) Auto-metric graph neural network based on a meta-learning strategy for the diagnosis of Alzheimer's disease," *IEEE Journal of Biomedical and Health Informatics*, 25(8):3141-3152. doi: 10.1109/JBHI.2021.3053568

Sun, Z., Yin, H., Chen, H., Chen, T., Cui, L. and Yang, F. (2021) Disease prediction via graph neural networks. *IEEE Journal of Biomedical and Health Informatics*, 25(3):818-826.

Vivar, G., Kazi, A., Burwinkel, H., Zwergal, A., Navab, N., Ahmadi, S.-A. (2021) Simultaneous imputation and classification using Multigraph Geometric Matrix Completion (MGMC): Application to neurodegenerative disease classification. *Artificial Intelligence in Medicine*, 117:102097. doi: 10.1016/j.artmed.2021.102097

Vogel, J.W., Young, A.L., Oxtoby, N.P. *et al.* (2021) Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nature Medicine*, 27:871–881 (2021). doi: 10.1038/s41591-021-01309-6

Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., Wang, C., Fu, H., Ma, Q. and Xu, D. (2021) scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature Communications,* 12:1882. doi: 10.1038/s41467-021-22197-x

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., Alzhemier's Disease Neuroimaging Initiative; Australian Imaging Biomarkers and Lifestyle flagship study of ageing (2020) Convolutional neural networks for classification of Alzheimer's disease: Overview and reproducible evaluation. *Medical Image Analysis,* 63:101694. doi: 10.1016/j.media.2020.101694

Wong-Lin, K., McClean, P.L., McCombe, N., Kaur, D., Sanchez-Bornot, J.M., Gillespie, P., Todd, S., Finn, D.P., Joshi, A., Kane, J. and McGuinness, B. (2020) Shaping a data-driven era in dementia care pathway through computational neurology approaches. *BMC Medicine*, 18(1):398. doi: 10.1186/s12916-020-01841-1

World Health Organisation (2020). Dementia: Key Facts, WHO.

Yang, K., Mohammed, E.A. and Far, B.H. (2021) Detection of Alzheimer's disease using graph-regularized convolutional neural network based on structural similarity learning of brain magnetic resonance images. *In: 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, 2021, pp. 326-333. doi:10.1109/IRI51335.2021.00051

Ying, R., Bourgeois, D., You, J., Zitnik, M. and Leskovec, J. (2019) GNNEexplainer: Generating explanations for graph neural networks. *In: Advances in Neural Information Processing Systems*, 32: 9240–9251.

Zhang, Z. (2016) Missing data imputation: focusing on single imputation. *Annals of Translational Medicine,* 4(1):9. doi: 10.3978/j.issn.2305-5839.2015.12.38.

Zhu, W. and Razavian, N. (2021) Variationally regularized graph-based representation learning for electronic health records. *arXiv preprints*, pp. 1-13, 2021. arXiv:1912.03761v2

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.