Multi-Modal Hypergraph Diffusion Network with Dual Prior for Alzheimer Classification

Angelica I. Aviles-Rivero\textsuperscript{1}, Christina Runkel\textsuperscript{1}, Nicolas Papadakis\textsuperscript{2}, Zoe Kourtzi\textsuperscript{3}, and Carola-Bibiane Schönlieb\textsuperscript{1}⋆

\textsuperscript{1} DAMTP, University of Cambridge, UK \{ai323,cr661,cbs31\}@cam.ac.uk
\textsuperscript{2} IMB, Universite de Bordeaux, France nicolas.papadakis@math.u-bordeaux.fr
\textsuperscript{3} Department of Psychology, University of Cambridge, UK zk240@cam.ac.uk

Abstract. The automatic early diagnosis of prodromal stages of Alzheimer’s disease is of great relevance for patient treatment to improve quality of life. We address this problem as a multi-modal classification task. Multi-modal data provides richer and complementary information. However, existing techniques only consider lower order relations between the data and single/multi-modal imaging data. In this work, we introduce a novel semi-supervised hypergraph learning framework for Alzheimer’s disease diagnosis. Our framework allows for higher-order relations among multi-modal imaging and non-imaging data whilst requiring a tiny labelled set. Firstly, we introduce a dual embedding strategy for constructing a robust hypergraph that preserves the data semantics. We achieve this by enforcing perturbation invariance at the image and graph levels using a contrastive based mechanism. Secondly, we present a dynamically adjusted hypergraph diffusion model, via a semi-explicit flow, to improve the predictive uncertainty. We demonstrate, through our experiments, that our framework is able to outperform current techniques for Alzheimer’s disease diagnosis.

Keywords: Hypergraph Learning · Multi-Modal Classification · Semi-Supervised Learning · Perturbation Invariance · Alzheimer’s disease.

1 Introduction

Alzheimer’s disease (AD) is an irreversible, neurodegenerative disease impairing memory, language and cognition \cite{3}. It starts slowly but progressively worsens and causes approximately 60-80\% of all cases of dementia \cite{4}. As there is no cure available yet, detecting the disease as early as possible, in prodromal stages, is crucial for slowing down its progression and for patient to improve quality of life. The body of literature has shown the potentials of existing machine learning methods e.g., \cite{25,21,26}. However, there are two major limitations of existing techniques. Firstly, incorporating other relevant data types such as phenotypic data in combination with large-scale imaging data has shown to be beneficial e.g. \cite{20}. However, existing approaches fail to exploit the rich available imaging

⋆ the Alzheimer’s Disease Neuroimaging Initiative
and non-imaging data. This is mainly due to the inherent problem of how to provide better and higher relations between multi-modal sources. Secondly, whilst vast amount of data is acquired every day at hospitals, labelling is expensive, time-consuming and prone to human bias. Therefore, to develop models that rely on extreme minimal supervision is of a great interest in the medical domain.

To address the aforementioned problems, hypergraph learning has been explored – as it allows going beyond pair-wise data relations. In particular, hypergraphs have already been explored for the task of AD diagnosis e.g. [19,31,24,25]. From the machine learning perspective [6], several works have addressed the problem of hypergraph learning, by generalising the graph Laplacian to hypergraphs e.g. [30,15,23]. The principles from such techniques opened the door to extend the hypergraph Laplacian to graph neural networks (GNNs) e.g. [27,11]. However, the commonality of existing hypergraph techniques for AD diagnosis, and in general in the ML community, is that they consider the clique/star expansion [2], and in particular, follow the hypergraph normalised cut of that [30].

To tackle the aforementioned challenges, we introduce a novel semi-supervised hypergraph learning framework for Alzheimer’s disease diagnosis. Our work follows a hybrid perspective, where we propose a new technique based on a dual embedding strategy and a diffusion model. To the best of our knowledge, this is the first work that explores invariance at the image and graph levels, and goes beyond the go-to technique of [30] by introducing a better hypergraph functional.

Our contributions are as follows. 1) We introduce a self-supervised dual multi-modal embedding strategy. The framework enforces invariance on two spaces – the manifold that lies the imaging data and the space of the hypergraph structure. Our dual strategy provides better priors on the data distribution, and therefore, we construct a robust graph that offers high generalisation capabilities. 2) In contrast to existing techniques that follow [30], we introduce a more robust diffusion-model. Our model is based on the Rayleigh quotient for hypergraph $p$-Laplacian and follows a semi-explicit flow.

2 Proposed Framework

This section introduces our semi-supervised hypergraph learning framework (see Fig. 1) highlighting two key parts: i) our dual embedding strategy to construct a robust graph, and ii) our dynamically adaptive hypergraph diffusion model.

2.1 Hypergraph Embeddings & Construction

The first part of our framework addresses a major challenge in hypergraph learning – that is, how to extract meaningful embeddings, from the different given modalities, to construct a robust hypergraph. The most recent works for Alzheimer’s disease diagnosis are based on extracting embeddings using sparse constraints or directly from a deep network. However, existing works only consider data level features to construct a graph or hypergraph and mainly for imaging data. In contrast to existing works, we allow for higher-order relations
Given:
Imaging Data
Comparison
Modality 1
Comparison
Modality m
Comparison
Classification Output
Uncertainty
Update
Diffusion
Model
Concatenate
Phenotypic Data

Fig. 1: Visual illustration of our framework. We first use a dual embeddings strategy at the image and hypergraph level. We then introduce a diffusion model for predicting early diagnosis of prodromal stages of Alzheimer’s disease.

between multi-modal imaging and non-imaging data. Secondly, we enforce a dual perturbation invariance strategy at both image and graph levels. With our dual multi-modal embedding strategy, we seek to provide better priors on the data distribution such as our model is invariant to perturbations, and therefore, we can construct a robust graph that offers high generalisation capabilities.

We consider a hypergraph as a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$, where $\mathcal{V} = \{v_1, ..., v_n\}$, $|\mathcal{V}| = n$ is a set of nodes and $\mathcal{E} = \{e_1, ..., e_m\}$. $|\mathcal{E}| = m$ the hyperedges. Moreover, $w : \mathcal{E} \to \mathbb{R}_{>0}$ refers to the hyperedges weights, in which each hyperedge $e$ is associated to a subset of nodes. The associated incidence matrix is given by $\mathcal{H}_{i,e} = \begin{cases} 1 & \text{if } i \in e, \\ 0 & \text{otherwise} \end{cases}$, where $i \in \mathcal{V}$ and $e$ is an hyperedge; i.e. a subset of nodes of $\mathcal{V}$. In our multi-modal setting, we assume a given $M$ modalities. We have for each modality $N$ samples $X = \{x_1, ..., x_N\} \in \mathcal{X}$ sampled from a probability distribution $\mathbb{P}$ on $\mathcal{X}$. We then have a multi-modal data collection as $\{X_1, ..., X_s, X_{s+1}, ..., X_M\}$. The first $s$ modalities are imaging data (e.g. PET imaging) whilst the remaining are non-imaging data (e.g. genetics).

The first problem we address in our framework is how to learn embeddings, $v = f_\theta (x)$, without supervision and invariant to perturbations. That is, given a deep network $f_\theta$, with parameters $\theta$, we seek to map the imaging/non-imaging sample $x$ to a feature space $v$. As we seek to obtain relevant priors on the data distribution, we also consider a group of $T$ transformations such as there exists a representation function $\phi : T \times \mathcal{X} \to \mathcal{X}$, $(t, x) \mapsto \phi(t, x)$. We then seek to enforce $\phi(t, x) = \phi(x)$ for all $t \in T$. We divide our embedding learning problem into two strategies corresponding to each type of data.
For the imaging data, we use contrastive self-supervised learning (e.g. [13, 14, 7]) for mapping $X$ to a feature space $v = \{v_1, ..., v_N\}$ with $v_i = f_\theta(x_i)$ such that $v_i$ better represents $x_i$. Given $t(x_i)$ and $t'(x_i)$, where $t, t' \sim T$ are operators that produce two perturbed versions of the given sample (defined following [8]), we seek to learn a batch of embeddings that are invariant to any transformation (see Figure 2). Formally, we compute the following contrastive loss:

$$L(i, j), X_i \in s_{\text{visual}} = -\log \frac{\exp(f_{i,j}/\tau)}{\sum_{k=1}^{N} \mathbb{1}_{[k \neq i]} \exp(f_{i,k}/\tau)},$$

where $\tau > 0$ is a temperature hyperparameter and $f_{i,k}$ follows the same cosine similarity definition than $f_{i,j}$. We denote the $k$-nearest neighbors of an embedding $v_i$ as $\text{NN}_k(v_i)$. We then construct for each $X_1, ..., X_s$ a corresponding hypergraph $H^{1}, ..., s_{\text{visual}}$. Given a set of phenotypic measures, we compute a $\text{NN}_k$ graph between subjects given the set of measures $z$ such as $S(x, z)$ if $x \in \text{NN}_k(z)$, otherwise 0; being $S$ a similarity function. That is, we enforce the connection of subjects within similar phenotypic measures. We then seek to perturb $H^{s+1,..., M}$ such that the transformed versions $\tilde{H}^{s+1,..., M} \sim q(\tilde{H}^{s+1,..., M}|H^{s+1,..., M})$ where $q(\cdot)$ refers to the transformation drawn from the distribution of the given measure. Our set of transformations are node dropping and edge perturbation. We follow, for the group of transformations $T$, the ratio and dropping probability strategies of [22, 29].

To maximise the agreement between the computed transformation, we use the same contrastive loss as defined in (1). We denote the loss as $L_{\text{pheno}}$ for non-imaging data. The final hypergraph, $\mathcal{H}$, is the result of concatenating all hypergraph structures from all modalities, $\mathcal{H}^1, ..., \mathcal{H}^M$ given by both imaging and non-imaging data.

### 2.2 Dynamically Adjusted Hypergraph Diffusion Model

The second key part of our framework is a semi-supervised hypergraph diffusion model. After constructing a robust hypergraph, we now detail how we perform disease classification using only a tiny labelled set.

We consider a small amount of labelled data $X_L = \{(x_i, y_i)\}_{i=1}^l$ with provided labels $L = \{1, .., L\}$ and $y_i \in L$ along with a large unlabelled set $X_u = \{x_k\}_{k=l+1}^m$. Fig. 2: We seek for distinctiveness. Related samples should have similar features.
The entire dataset is then $X = X_L \cup X_U$. The goal is to use the provided labeled data $X_L$ to infer a function, $f : \mathcal{X} \mapsto \mathcal{Y}$, that maps the unlabelled set to class labels. However, to obtain such mapping efficiently and with minimum generalisation error is even more challenging when using multi-modal data e.g. [20]. We cast this problem as a label diffusion process on hypergraphs.

We consider the problem of hypergraph regularisation for semi-supervised classification [30][15][11], where we aim to find a function $u^*$ to infer labels for the unlabelled set and enforce smoothness on the hypergraph. We seek to solve a problem of the form $u^* = \arg\min_{u \in \mathcal{R}^n} \{\mathcal{L}_{\text{emp}}(u) + \gamma \mathcal{O}(u)\}$, where $u \mathcal{L}_{\text{emp}}(u)$ refers to an empirical loss (e.g., square loss), $\mathcal{O}(u)$ denotes a regulariser and $\gamma > 0$ is a weighting parameter balancing the importance of each term. Whilst majority of works consider the clique/star expansion [2] and follow the principles of [30], we seek to provide a more robust functional to avoid the inherent disadvantages of such approximations (e.g. bias during the graph diffusion [15]). In particular, we consider the setting of Rayleigh quotient for $p$-Laplacian. We then seek to estimate, for $p = 1$, the particular family of solutions based on the minimisation of the ratio $\Omega(u) = \frac{\text{TV}_H(u)}{|u|}$, where $\text{TV}_H = \sum_{e \in \mathcal{E}} w_e \max_{i, j \in e} |u_i - u_j|$ is the total variation functional on the hypergraph. To do this, we generalise [10] to a hypergraph setting and introduce a dynamic adjustment on the diffusion model, which is based on controlling the predictive uncertainty. Our framework allows for both binary and multi-class classification settings. Given a number of epochs $E \in [1, ..., E]$ and following previous notation, we compute an alternating optimisation process as follows.

**Alternating Optimisation for Epoch $\in [1, ..., E]$**:

**Sub-Problem 1: Diffusion Model Minimisation.** For a given label, we define a function $u \in \mathcal{R}^n$ over the nodes, and denote its value at node $x$ as $u(x)$. In this binary setting, the objective is to estimate a function $u$ that denotes the presence (resp. absence) of the label for data $x$. To that end, following [10], we seek to solve $u^*$ to guide the partitioning and estimate error is even more challenging when using multi-modal data e.g. [20].

The entire dataset is then $X = X_L \cup X_U$. The goal is to use the provided labeled data $X_L$ to infer a function, $f : \mathcal{X} \mapsto \mathcal{Y}$, that maps the unlabelled set to class labels. However, to obtain such mapping efficiently and with minimum generalisation error is even more challenging when using multi-modal data e.g. [20]. We cast this problem as a label diffusion process on hypergraphs.

We consider the problem of hypergraph regularisation for semi-supervised classification [30][15][11], where we aim to find a function $u^*$ to infer labels for the unlabelled set and enforce smoothness on the hypergraph. We seek to solve a problem of the form $u^* = \arg\min_{u \in \mathcal{R}^n} \{\mathcal{L}_{\text{emp}}(u) + \gamma \mathcal{O}(u)\}$, where $u \mathcal{L}_{\text{emp}}(u)$ refers to an empirical loss (e.g., square loss), $\mathcal{O}(u)$ denotes a regulariser and $\gamma > 0$ is a weighting parameter balancing the importance of each term. Whilst majority of works consider the clique/star expansion [2] and follow the principles of [30], we seek to provide a more robust functional to avoid the inherent disadvantages of such approximations (e.g. bias during the graph diffusion [15]). In particular, we consider the setting of Rayleigh quotient for $p$-Laplacian. We then seek to estimate, for $p = 1$, the particular family of solutions based on the minimisation of the ratio $\Omega(u) = \frac{\text{TV}_H(u)}{|u|}$, where $\text{TV}_H = \sum_{e \in \mathcal{E}} w_e \max_{i, j \in e} |u_i - u_j|$ is the total variation functional on the hypergraph. To do this, we generalise [10] to a hypergraph setting and introduce a dynamic adjustment on the diffusion model, which is based on controlling the predictive uncertainty. Our framework allows for both binary and multi-class classification settings. Given a number of epochs $E \in [1, ..., E]$ and following previous notation, we compute an alternating optimisation process as follows.

**Alternating Optimisation for Epoch $\in [1, ..., E]$**:

**Sub-Problem 1: Diffusion Model Minimisation.** For a given label, we define a function $u \in \mathcal{R}^n$ over the nodes, and denote its value at node $x$ as $u(x)$. In this binary setting, the objective is to estimate a function $u$ that denotes the presence (resp. absence) of the label for data $x$. To that end, following [10], we seek to solve $u^*$ to guide the partitioning and estimate error is even more challenging when using multi-modal data e.g. [20].
(e.g. [3]) for the unlabelled set, as it is conditioned solely to the tiny labelled set as prior. We then seek to update the uncertainty of the computed \( \hat{y}_i \) via solving Sub-Problem 2.

**Sub-Problem 2: Uncertainty Hypergraph Minimisation.** Following the scheme from Figure 1 and notation from previous sections, we first initialise \( f_\psi \) using the tiny labelled set available, \( X_L \), and then we take \( \hat{y}_i \) from Sub-Problem 1 and check the uncertainty of \( \hat{y}_i \) by computing the following loss:

\[
\mathcal{L}_{DYN}(X, Y, \hat{Y}; \theta) := \min_{\theta} \sum_{i=1}^{l} \mathcal{L}_{CE}(f_\theta(x_i), y_i) + \sum_{i=l+1}^{m} \gamma_i \mathcal{L}_{CE}(f_\theta(x_i), \hat{y}_i),
\]

where \( \mathcal{L}_{CE} \) is a cross entropy loss and \( \gamma \) is the measure of uncertainty via entropy [18,1] defined as \( \gamma_i = 1 - (H(u_{\hat{y}_i})/\log(L)) \), where \( H \) refers to the entropy and \( u_{\hat{y}_i} \) is normalised beforehand. **Intuition:** From previous step, we obtain an initial prediction for the unlabelled set on the hypergraph. However, in semi-supervised learning due to the tiny labelled set as prior, the unlabelled data tend to have an asymptotic bias in the prediction. We then ensure that there is a high certainty in the unlabelled predictions since early epochs to avoid the propagation of incorrect predictions.

3 Experimental Results

In this section, we provide all details of our evaluation protocol.

**Data Description.** We evaluate our semi-supervised hypergraph framework using the Alzheimer’s disease Neuroimaging Initiative (ADNI) dataset[^1]. ADNI is a multi-centre dataset composed of multi-modal data including imaging and multiple phenotype data. We consider 500 patients using MRI, PET, demographics and Apolipoprotein E (APOE). We included APOE as it is known to be a crucial genetic risk factor for developing Alzheimer’s disease. The dataset contains four categories: early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), normal control (NC) and Alzheimer’s disease (AD).

**Evaluation Protocol.** We design a three part evaluation scheme. Firstly, we follow the majority of techniques convention for binary classification comparing AD vs NC, AD vs EMCI, AD vs LMCI, EMCI vs NC, LMCI vs NC and EMCI vs LMCI (see Supplementary Material for extended results). Secondly, we extended the classification problem to a multi-class setting including the four classes AD vs NC vs EMCI vs LMCI. We consider this setting, as one of the major challenges in AD diagnosis is to fully automate the task without pre-selecting classes [12].

To evaluate our model, we performed comparisons with state-of-the-art techniques on hypergraph learning: HF [24], HGSCCA [25], HGNN [11] and DHGNN [16]. We also added the comparison against GNNs [20], which is based on graph neural networks (GNNs). We added this comparison as it also considers imaging

[^1]: Data used were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)
Technique | AD vs NC | EMCI vs LMCI | LMCI vs NC | Ours
---|---|---|---|---
GNNs [20] | 81.60±2.81 | 83.20±3.10 | 80.62±2.30 | 83.20±2.10 | 80.62±2.30 | 84.01±2.16
HF [24] | 87.20±3.10 | 88.01±2.15 | 86.60±2.60 | 87.20±3.10 | 88.01±2.15 | 89.20±2.70
HGSCCA [25] | 85.60±2.16 | 86.20±3.11 | 84.40±2.15 | 85.60±2.16 | 86.20±3.11 | 86.80±2.70
HGNN [11] | 88.01±2.60 | 90.40±2.16 | 87.59±2.42 | 88.01±2.60 | 90.40±2.16 | 90.20±2.45
DHGNN [16] | 89.90±2.40 | 89.60±2.15 | 90.21±2.45 | 89.90±2.40 | 89.60±2.15 | 90.00±2.55
Ours | 92.11±2.03 | 92.80±2.16 | 91.33±2.43 | 92.11±2.03 | 92.80±2.16 | 92.00±2.45

Table 1: Numerical comparison of our technique and existing (graph) and hypergraph techniques. All comparison are run on same conditions. The best results are highlighted in green colour.

Technique | LMCI vs NC | Ours
---|---|---
GNNs [20] | 72.40±2.05 | 82.01±2.16
HF [24] | 77.01±2.26 | 84.01±2.34
HGSCCA [25] | 74.00±2.10 | 82.01±2.16
HGNN [11] | 78.90±3.01 | 81.80±2.55
DHGNN [16] | 79.20±2.70 | 81.80±2.55
Ours | 82.01±2.16 | 84.01±2.34

Table 2: Performance comparison of our technique and existing hypergraph models for LMCI vs NC. The results in green colour denotes the highest performance.

Results & Discussion. We began by evaluating our framework following the binary comparison of AD vs NC, EMCI vs LMCI and LMCI vs NC (see supplementary material for extended results). We report a detailed quantitative analysis to understand the performance of ours and compared techniques. We use 15% of...
Table 3: Error rate comparison of our technique against existing models. The results in green denotes the best performance.

| Technique | AD vs NC vs MCI Error Rate & 95%CI | AD vs NC vs EMCI vs LMCI Error Rate & 95%CI |
|-----------|-----------------------------------|------------------------------------------|
| GNNs [20] | 36.19±4.45                        | 39.01±3.12                               |
| HGNN [11] | 26.35±3.20                        | 28.09±3.65                               |
| DHGNN [16] | 23.10±2.60                       | 26.25±2.55                               |
| Ours      | 16.25±2.22                        | 18.31±2.45                               |

Fig. 4: (left side) Ablation study for the components of our technique. (right side) Performance comparison using different type of multi-modal data.

labelled data (see supplementary material). The results are reported in Tables 1 & 2. In a closer look at the results, we observe that our technique reports the best performance for all comparison and in all metrics. We also observe that the GNNs [20] reported the worst performance. This is due to the fact that GNNs does not allow for higher-order relations on the graph. We highlight that whilst the other techniques reported a good performance, our technique substantially improved over all methods (with statistical significance, see supplementary material). We underline that all compared techniques follows the approximation of [30], which introduces bias during the hypergraph partition. We showed that our diffusion model, that follows different principles, mitigates that problem.

To further support our previous results, we also report results in the more challenging setting of multi-class classification for AD diagnosis. The results are reported in Table 3. We observe that our technique performance is consistent with previous results. That is, our technique reported the lowest error rate for both multi-class cases. Our technique decreased the error rate for more than 40% for all techniques. Figure 3 complements these results by reporting the performance with respect to the four classes case.

Finally, to support the design of our technique, we performed two ablations studies regarding our design and the modalities used. We start by evaluating the influence of our full model against without our dual perturbed embedding strategy (displayed as DP) and our uncertainty (denoted as UC) scheme. The results are reported at the left side of Figure 4 on the four classes case. Whilst our dual strategy substantially improves the performance, we observe that our uncertainty scheme has a major repercussion. The intuition is that our diffusion model ensures, at early epochs, that there is a high certainty in the prediction
avoiding incorrect prediction in subsequent epochs. We also include an ablation study regarding the impact of the modalities (PT refers to phenotypic data). From the right side of Figure \[4\] we observe that including phenotypic data has a greater positive impact on the performance than using only imaging data.

4 Conclusion

We proposed a novel semi-supervised hypergraph framework for Alzheimer’s disease diagnosis. Unlike existing techniques, we introduce a dual embedding strategy. As phenotypic data highly differs from imaging data. Moreover, in contrast to existing techniques that follow the hypergraph approximation of \[30\]. We introduce a better diffusion model, which solution is provided by a semi-explicit flow in the fra hypergraph learning. From our results, we showed that our technique outperforms other hypergraph techniques. Future work includes a more extensive clinical evaluation with public and in-home datasets.

Acknowledgements AIAR acknowledges support from CMIH and CCIMI, University of Cambridge. CR acknowledges support from the CCIMI and the EPSRC grant EP/W524141/1 ref. 2602161. ZK acknowledges support from the BBSRC (H012508, BB/P021255/1), Wellcome Trust (205067/Z/16/Z, 221633/Z/20/Z) and Royal Society (INF/R2/202107). CBS acknowledges the Philip Leverhulme Prize, the EPSRC fellowship EP/V029428/1, EPSRC grants EP/T003553/1, EP/N014588/1, Wellcome Trust 215733/Z/19/Z and 221633/Z/20/Z, Horizon 2020 No. 777826 NoMADS and the CCIMI.

References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion 76, 243–297 (2021)
2. Agarwal, S., Branson, K., Belongie, S.: Higher order learning with graphs. In: Proceedings of the 23rd international conference on Machine learning (2006)
3. Arazo, E., Ortego, D., Albert, P., O’Connor, N.E., McGuinness, K.: Pseudo-labeling and confirmation bias in deep semi-supervised learning. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2020)
4. Association, A.: What is alzheimer’s disease? (2022)
5. Aviles-Rivero, A.I., Papadakis, N., Li, R., Sellars, P., Fan, Q., Tan, R.T., Schönlieb, C.B.: GraphXnet - chest x-ray classification under extreme minimal supervision. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 504–512. Springer (2019)
6. Berge, C.: Hypergraphs: combinatorics of finite sets, vol. 45. Elsevier (1984)
7. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
8. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 702–703 (2020)
9. De Strooper, B., Karran, E.: The cellular phase of alzheimer’s disease. Cell 164(4), 603–615 (2016)
10. Feld, T., Aujol, J.F., Gilboa, G., Papadakis, N.: Rayleigh quotient minimization for absolutely one-homogeneous functionals. Inverse Problems 35(6), 064003 (2019)
11. Feng, Y., You, H., Zhang, Z., Ji, R., Gao, Y.: Hypergraph neural networks. In: AAAI Conference on Artificial Intelligence. vol. 33, pp. 3558–3565 (2019)
12. Goenka, N., Tiwari, S.: Deep learning for alzheimer prediction using brain biomarkers. Artificial Intelligence Review 54(7), 4827–4871 (2021)
13. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06). vol. 2, pp. 1735–1742. IEEE (2006)
14. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9729–9738 (2020)
15. Hein, M., Setzer, S., Jost, L., Rangapuram, S.S.: The total variation on hypergraphs-learning on hypergraphs revisited. Advances in Neural Information Processing Systems 26 (2013)
16. Jiang, J., Wei, Y., Feng, Y., Cao, J., Gao, Y.: Dynamic hypergraph neural networks. In: IJCAI. pp. 2635–2641 (2019)
17. Joachims, T., et al.: Transductive inference for text classification using support vector machines. In: ICML. vol. 99, pp. 200–209 (1999)
18. Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? Advances in neural information processing systems (2017)
19. Pan, J., Lei, B., Shen, Y., Liu, Y., Feng, Z., Wang, S.: Characterization multimodal connectivity of brain network by hypergraph gan for alzheimer’s disease analysis. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 467–478. Springer (2021)
20. Parisot, S., Ktena, S.I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., Rueckert, D.: Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer’s disease. Medical Image Analysis (2018)
21. Pölsler, S., Aigner, C., Wachinger, C.: Scalable, axiomatic explanations of deep alzheimer’s diagnosis from heterogeneous data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (2021)
22. Qin, J., Chen, Q., Dong, Y., Zhang, J., Yang, H., Ding, M., Wang, K., Tang, J.: Gcc: Graph contrastive coding for graph neural network pre-training. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1150–1160 (2020)
23. Saito, S., Mandic, D., Suzuki, H.: Hypergraph p-laplacian: A differential geometry view. In: AAAI Conference on Artificial Intelligence. vol. 32 (2018)
24. Shao, W., Peng, Y., Zu, C., Wang, M., Zhang, D., Initiative, A.D.N., et al.: Hypergraph based multi-task feature selection for multimodal classification of alzheimer’s disease. Computerized Medical Imaging and Graphics 80, 101663 (2020)
25. Shao, W., Xiang, S., Zhang, Z., Huang, K., Zhang, J.: Hyper-graph based sparse canonical correlation analysis for the diagnosis of alzheimer’s disease from multi-dimensional genomic data. Methods 189, 86–94 (2021)
26. Shin, H.C., Ihsani, A., Xu, Z., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J., Initiative, A.D.N., et al.: Gandalf: Generative adversarial networks with
discriminator-adaptive loss fine-tuning for alzheimer’s disease diagnosis from mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 688–697. Springer (2020)

27. Yadati, N., Nimishakavi, M., Yadav, P., Nitin, V., Louis, A., Talukdar, P.: Hypergcn: A new method for training graph convolutional networks on hypergraphs. Advances in neural information processing systems 32 (2019)

28. Yang, F., Meng, R., Cho, H., Wu, G., Kim, W.H.: Disentangled sequential graph autoencoder for preclinical alzheimer’s disease characterizations from adni study. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 362–372. Springer (2021)

29. You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., Shen, Y.: Graph contrastive learning with augmentations. Advances in Neural Information Processing Systems 33, 5812–5823 (2020)

30. Zhou, D., Huang, J., Schölkopf, B.: Learning with hypergraphs: Clustering, classification, and embedding. Advances in neural information processing systems (2006)

31. Zuo, Q., Lei, B., Shen, Y., Liu, Y., Feng, Z., Wang, S.: Multimodal representations learning and adversarial hypergraph fusion for early alzheimer’s disease prediction. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 479–490. Springer (2021)
Supplemental Material

Multi-Modal Hypergraph Diffusion Network with Dual Prior for Alzheimer Classification

Angelica I. Aviles-Rivero, Christina Runkel, Nicolas Papadakis, Zoe Kourtzi and Carola-Bibiane Schonlieb

1DAMTP, University of Cambridge, UK {ai323, cr661, cbs31}@cam.ac.uk
2IMB, Universite de Bordeaux, France nicolas.papadakis@math.u-bordeaux.fr
3Department of Psychology, University of Cambridge, UK zk240@cam.ac.uk

Supplementary Numerical Results & Details

A. Supplementary Results. In this section, we extend the numerical results of Table 1 from the main paper. Our results report the performance of AD vs EMCI, AD vs LMCI and NC vs EMCI. The comparison includes GNNs [20], HF [24], HGSCCA [25], HGNN [11] and DHGNN [16]. Following standard protocol in semi-supervised learning, we report the results over five repeated randomly selected labelled sets, and report the mean of metrics along with the standard deviation. The results are reported in Fig. 5. In a closer look at the results, we observe that the performance behaviour of our technique is in line to the reported in the main paper. That is, our technique consistently outperforms by a significant margin the compared techniques up to ∼ 19% with respect to the GNNs and in the range of ∼ 8 to ∼ 16% compared with other hypergraph techniques.

Fig. 5: Performance comparison of our technique vs existing techniques. All results are run under the same conditions.

B. Statistical Tests. To further support our technique, we performed a set of non-parametric statistical tests. For the binary comparisons, we performed a
Wilcoxon test, and we found that the difference in performance of our technique vs the compared ones is statistically significant \((p < 0.0001)\). For the multi-class setting (4-classes case), we use the non-parametric Friedman test, followed by a pair-wise comparison with adjusted \(p\)-values using the Bonferroni method. We found that our technique was statistically significant, \((\chi^2 = 18.1, p = 0.0002)\), different with respect to all compared techniques. Hence, the generalisation and robustness of our technique is demonstrated.

\[\text{C. Supplementary Details.}\] We used ADNI-2 in our experiments as it provides an explicit separation between EMCI and LMCI. Moreover, for the imaging data we used Florbetapir (FBP) PET and T1 MR images. Our experiments from the main paper were computed using only 15\% of labelled data. However, we provide an ablation study regarding the behaviour of our technique under different labelled counts. The results are reported for the four classes case, and in terms of error rate. The results are displayed in Fig. 6. We use 15\% of labelled data for the following reasons. Firstly, we want to use a labelled rate as lower as possible. Secondly, as displayed in our results 15\% is a good trade-off between performance and label rate. We can observe that labelled rates greater than 25\% only provides a tiny improvement in performance. This performance effect is well-known in semi-supervised learning, as the transductive effect e.g. [17], and also observed in our experiments.