Self-pruning Graph Neural Network for Predicting Inflammatory Disease Activity in Multiple Sclerosis from Brain MR Images

Chinmay Prabhakar[1(✉)], Hongwei Bran Li[1,2,5], Johannes C. Paetzold[3], Timo Loehr[2], Chen Niu[4], Mark Mühlau[5], Daniel Rueckert[2,3], Benedikt Wiestler[5], and Bjoern Menze[1,2]

[1] Department of Quantitative Biomedicine, University of Zurich, Zürich, Switzerland
chinmay.prabhakar@uzh.ch
[2] Department of Computer Science, Technical University of Munich, Munich, Germany
[3] Department of Computing, Imperial College London, London, UK
[4] The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China
[5] Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

Abstract. Multiple Sclerosis (MS) is a severe neurological disease characterized by inflammatory lesions in the central nervous system. Hence, predicting inflammatory disease activity is crucial for disease assessment and treatment. However, MS lesions can occur throughout the brain and vary in shape, size and total count among patients. The high variance in lesion load and locations makes it challenging for machine learning methods to learn a globally effective representation of whole-brain MRI scans to assess and predict disease. Technically it is non-trivial to incorporate essential biomarkers such as lesion load or spatial proximity. Our work represents the first attempt to utilize graph neural networks (GNN) to aggregate these biomarkers for a novel global representation. We propose a two-stage MS inflammatory disease activity prediction approach. First, a 3D segmentation network detects lesions, and a self-supervised algorithm extracts their image features. Second, the detected lesions are used to build a patient graph. The lesions act as nodes in the graph and are initialized with image features extracted in the first stage. Finally, the lesions are connected based on their spatial proximity and the inflammatory disease activity prediction is formulated as a graph classification task. Furthermore, we propose a self-pruning strategy to auto-select the most critical lesions for prediction. Our proposed method outperforms the existing baseline by a large margin (AUCs of 0.67 *vs.* 0.61 and 0.66 *vs.* 0.60 for one-year and two-year inflammatory disease activity, respectively). Finally, our proposed method enjoys inherent explainability by

B. Wiestler and B. Menze—Contributed equally as senior authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43993-3_22.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14227, pp. 226–236, 2023.
https://doi.org/10.1007/978-3-031-43993-3_22

assigning an importance score to each lesion for the overall prediction. Code is available at https://github.com/chinmay5/ms_ida.git.

1 Introduction

Multiple Sclerosis (MS) is a severe central nervous system disease with a highly nonlinear disease course where periodic relapses impair the patient's quality of life. Clinical studies show that relapses co-occur with the appearance of new inflammatory MS lesions in MR images [19,25], making MR imaging a central element for the clinical management of MS patients. Further, assessing new MS lesions is crucial for disease assessment and therapy monitoring [5,8]. Unfortunately, prevailing therapies often involve highly active immunomodulatory drugs with potentially severe side effects. Hence, it necessitates developing machine learning models capable of predicting the future disease activity of individual patients to select the best therapy.

While recent approaches applied convolutional neural networks (CNN) to directly learn features from MR image space [3,4,21,26,28], there remain challenges to obtain an effective global representation to characterize disease status. We attribute the difficulty of MS inflammatory disease activity prediction to a set of distinct disease characteristics that are well observable in MR images. First, while lesions have a sufficient signal-to-noise profile in images, their variation in shape, size, and number of occurrences amongst patients make it challenging for existing CNN-based methods that process the whole MRI scan in one go. Second, with advanced age, small areas appearing in the MRI of healthy individuals may resemble MS. As such, it is crucial to not only predict MS but also to identify the lesions deemed consequential for the final prediction.

To solve this problem, we use concepts from geometric deep learning. Specifically, we propose a two-stage pipeline. First, the lesions in the MRI scans are segmented using a state-of-the-art 3D segmentation algorithm [9], and their image features are extracted with a self-supervised method [14]. Second, the extracted lesions are converted into a patient graph. The lesions act as nodes of the graph, while the edge connectivity is determined using the spatial proximity of the lesions. By this representation, we can solve the MS inflammatory disease activity prediction task as a graph-level classification problem. We argue that formulating the MS inflammatory disease activity prediction in our two-stage pipeline has the following advantages: (1) Graph neural networks can easily handle different numbers of nodes (lesions) and efficiently incorporate their spatial locations. (2) Modern segmentation [9,16,23] and representation learning methods [2,11,14] are effective tools for lesion detection and allow us to extract pathology-specific features. (3) By operating at the lesion level, it is possible to discover the lesions that contribute most to the eventual prediction, making the decisions more interpretable. Thus, our proposed solution can be a viable methodology for MS inflammatory disease activity prediction to handle the associated challenges.

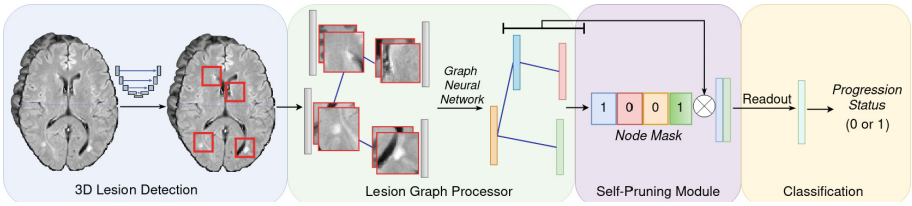

Fig. 1. Our proposed Multiple Sclerosis (MS) inflammatory disease activity prediction framework. We first detect lesions in the MRI scan using nn-Unet [9]. A crop centered at the detection is extracted and used to learn self-supervised lesion features. Next, we build a graph from these detected lesions, where each lesion becomes a node, with the connections (edges) between lesions (nodes) defined by spatial proximity. This graph is processed using a graph neural network to generate enriched lesion features. Next, our self-pruning module (SPM) processes these enriched lesion features to determine an importance score for each lesion. The least scoring lesions are pruned-off and the highest scoring lesions are passed to the readout layer to obtain a graph-level feature vector. This graph-level vector is used for the final prediction.

Contributions. Our contribution is threefold: (1) we are the first to formulate the MS inflammatory disease activity prediction task as a graph classification problem, thereby bringing a new set of methods to a significant clinical problem. (2) We propose a two-stage pipeline that effectively captures inherent MS variations in MRI scans, thus generating an effective global representation. (3) We develop a self-pruning module, which assigns an importance score to each lesion and reduces the task complexity by prioritizing the critical lesions. Additionally, the assigned per-lesion importance score improves our model's explainability.

2 Methodology

Overview. The objective is to predict MS inflammatory disease activity, i.e., to classify if new or significantly enlarged inflammatory lesions appear in the follow-up after the initial MRI scans. We denote the dataset as $D(X, y)$, where X is the set of lesion patches extracted from MR scans, and y $\in \{0, 1\}$ is the inflammatory disease activity status. For patient i, multiple lesion patches $\{x_1^i, x_2^i, ...x_n^i\}$ can exist, where n is the total number of lesions. We aim to learn a mapping function f: $\{x_1^i, x_2^i, ...x_n^i\} \rightarrow \{0, 1\}$. Please note that our formulation *differs* from existing methods [21,26,28], which aim to learn a direct mapping from the MR image to the inflammatory disease activity label. As shown in Fig. 1, our proposed method consists of four distinct components. We describe each component in the following sections.

Lesion Detection and Feature Extraction. We focus on the individual lesions instead of processing whole-brain MRI scans. This is important because MS lesions typically comprise less than 1% of voxels in the MRI scan. With this strategy, the graph model can aggregate lesion-level features for an effective patient-specific representation. First, we detect the lesions using a state-of-the-art nn-Unet [9] pre-trained with MR images and their consensus annotations

from two neuro-radiologists. Then, for each detected lesion, we extract a small fixed-size patch centered at it. Finally, we learn self-supervised features z for the lesion using a recent Transformer-based approach [14].

Lesion Graph Processor. In the second stage, we generate a patient-specific graph $G(V, E, Z)$ from the detected lesions. The lesions act as vertices V of this graph and are initialized with the crop-derived features Z. The spatial location s of the lesions is used to determine their connectivity E using a k-Nearest Neighbor (kNN) graph method [27]. Furthermore, the connected edges are weighted based on their spatial proximity. Specifically, given two lesions v_i and v_j, with spatial locations s_i and s_j respectively, the edge weight $w(v_i, v_j) = exp(-\frac{\|s_i - s_j\|^2}{\tau^2})$. τ is a scalar that controls the contribution of distant lesions. Hence the final graph connectivity can be represented as:

$$E_{i,j} = \begin{cases} w(v_i, v_j), \text{ if } j \in N(v_i) \text{ or } i \in N(v_j) \\ 0, \qquad\qquad \text{otherwise} \end{cases} \tag{1}$$

where $N(v_i)$ are the nodes directly connected to the node v_i. Constructing a graph from the lesions is instrumental in two aspects: (1) The framework allows us to work with varying numbers of lesions in different patients. Alternatively, sequential models could be employed. However, since the lesions lack a canonical ordering, such models would not achieve an effective global representation [12]. (2) It is possible to incorporate meaningful lesion properties such as spatial proximity and the number of lesions.

Please note that separate graphs are created for individual patients. Thus, MS inflammatory disease activity prediction is formulated as a graph-level classification task. The graph $G(V, E, Z)$ can be processed using message-passing neural networks (such as GCN [10], GAT [1], EdgeConv [24], GraphSage [6], to name a few) to learn enriched lesion features $\hat{Z}$. These enriched features are passed through to the self-pruning module.

Self-pruning Module. The number of lesions can vary substantially among the patients, including the possibility of false positives in the segmentation stage. As such, it is crucial to recognize the most relevant lesions for the final prediction. In addition, this will bring inherent explainability and make it easier for a doctor to validate model predictions. To accomplish this, the enriched lesion features $\hat{Z}$ are passed through a self-pruning module (SPM). The SPM produces a binary mask M for each lesion to determine whether a lesion contributes to the classification. The SPM uses a learnable projection vector $\vec{p}$ to compute importance scores $(\hat{Z}\vec{p}/||\vec{p}||)$ for the lesions. These scores are scaled with a sigmoid layer. We retain the high-scoring lesions and discard the rest, which is formulated as:

$$M = top\text{-}r(\sigma(\frac{\hat{Z}\vec{p}}{||\vec{p}||}), r) \tag{2}$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid function and $top\text{-}r(\cdot)$ is an operator which selects a fraction r of the lesions based on high importance score. r is a

hyper-parameter in our setup. Since the masking process is part of the forward pass through the model during both training and inference stages and not a post hoc modification, we refer to it as *self-pruning* of nodes. Features of the remaining nodes ($\hat{Z}' = \hat{Z} \otimes M$) are passed to the classification head.

It should be noted that the existence of multiple lesions is a typical characteristic of MS. Hence, a crucial aspect of MS management is that clinicians must identify signs of inflammatory disease activity in MR images to make treatment decisions [19]. Therefore, along with predicting inflammatory disease activity, interpreting the contribution of individual lesions to the prediction is essential. By assigning an importance score to each lesion, the SPM can provide explainability to clinicians at a lesion level, while existing CNN methods can not.

Classification Head. The classification head consists of a readout layer aggregating all the remaining node's features to produce a single feature vector $\hat{z}'$ for the entire graph. This graph-level feature is passed through an MLP to obtain the final prediction $\hat{y}$. We train our model using a binary cross-entropy loss.

$$\mathcal{L}_{clf} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(\hat{y_i}) + (1 - y_i)\log(1 - \hat{y_i})\right) \tag{3}$$

where N is the number of patients, y_i is the ground truth inflammatory disease activity information and $\hat{y_i}$ is the model prediction.

3 Experiments

Datasets and Image Preprocessing. Our approach is evaluated on a cohort of 430 MS patients collected following approval from the local IRB [7]. Patients included in this analysis were diagnosed with relapsing-remitting MS, with a maximum disease duration of three years at the time of baseline scan. We collect the FLAIR and T1w MR scans for each patient. The scans have a uniform voxel size of $1 \times 1 \times 1\,\mathrm{mm}^3$, were rigidly co-registered to the MNI152 atlas and skull-stripped using HD-BET [17]. Three neuro-radiologists independently read longitudinal subtraction imaging, where FLAIR images from two time points were co-registered and subtracted. In this vein, new and significantly enlarged lesions are identified as positive inflammatory disease activity.

The dataset contains MS inflammatory disease activity information for clinically relevant one-year and two-year intervals [20]. At the end of the first year, we have the inflammatory disease activity status of 430 patients, with 303 showing activity and 127 not. Similarly, at the end of two years, we have data available for 347 patients, with 287 showing activity and 60 not. Thus, the dataset shows a slight imbalance in favor of inflammatory disease activity. This imbalance is a typical property in the MS patients cohort that impairs algorithm development.

Feature Extraction and Training Configuration. We use an nn-Unet [9] for lesion segmentation and detection. Then a uniform crop of size $24 \times 24 \times 24$ mm^3 is extracted centered at each lesion. The cropped patches are passed through a transformer-based masked autoencoder [14] to extract self-supervised lesion features. The encoder produces a 768-dimensional feature vector for each patch. We also append normalized lesion coordinates to the encoder output to get the final lesion features.

The lesions are connected using a k-nearest neighbor algorithm with $k = 5$. Further, these connections (edges) are weighted using $\tau = 0.01$ (Eq. 1). Two message-passing layers with hidden dimensions of 64 and 8, respectively, process the generated graph to enrich lesion features. Next, the enriched features are passed through the SPM. The SPM uses a learnable projection vector $\vec{p} \in R^8$ and sigmoid activation to learn the importance score. Based on this importance score, a mask is produced to select $r = 0.5$ (i.e., 50%) of the highest-scoring lesions and discard the rest. Next, a sum aggregation is used as the readout function. These aggregated features are passed through 2 feed-forward layers with hidden dimensions of size 8. Finally, the features are passed to a sigmoid function to obtain the final prediction.

The model is trained for 300 epochs using *AdamW* optimizer [13] with 0.0001 weight decay. The base learning rate is 1e-4. The batch size is set to 16. A dropout layer with p = 0.5 is used between different feed-forward blocks. Since the dataset is imbalanced in favor of patients experiencing inflammatory disease activity, we use a balanced batch sampler to load approximately the same number of positive and negative samples in each mini-batch.

Evaluation Strategy, Classifier, and Metrics. We report our results on MS inflammatory disease activity prediction for the clinically relevant one and two-year intervals [20]. The Area Under the Receiver Operating Characteristic Curve (AUC) is used as the evaluation metric. We use 80% of the samples as the training set and 10% as the validation. The validation set is used to select the best model which is then applied to the remaining 10% cases. This procedure is iterated until all cases have been assigned to a test set once (ten-fold cross-validation). The same folds are used for the proposed model and baseline algorithms.

4 Results

Quantitative Comparison. Table 1 shows the classification performance for the ten folds on one-year and two-year lesion inflammatory disease activity prediction. The ± indicates the corresponding standard deviations. We compare our method against two existing approaches for MS inflammatory disease activity prediction baselines, a 3D Res-Net [28], and a multi-resolution CNN architecture [21]. These methods learn a direct mapping from the MR image to the inflammatory disease activity label. Our graph model outperforms the baseline methods on one (**0.67** *vs.* 0.61 AUC) and two-year inflammatory disease activity prediction (**0.66** *vs.* 0.60 AUC). In the following, we analyze and discuss each component of our established framework.

Table 1. Comparison of our method against the existing CNN-based solutions. We report the AUC score for MS Inflammatory Disease Activity (IDA) prediction at the end of one and two years. Our proposed two-stage solution outperforms the existing baselines, achieving the best AUC score on both prediction tasks.

Methods	One year IDA	Two year IDA
	AUC	AUC
3D ResNet	0.595 ± 0.104	0.575 ± 0.099
CNN multi-res [21]	0.610 ± 0.053	0.600 ± 0.059
Ours	**0.671 ± 0.062**	**0.664 ± 0.063**

Ablation Study. In this section, we analyze the importance of different components of our proposed method. We defer the analysis of the lesion feature representation to the appendix owing to space constraints.

The Effectiveness of Graph Structure. Since the lesion feature extractor generates rich lesion features, one may argue that the graph structure is unwarranted. There are two alternatives to using a graph, (i) completely discard the graph structure, use a feed-forward layer to enrich the lesion features further, and aggregate them to perform classification [15] (Since this formulation regards the input as a set, we call this *Set-Proc* model). (ii) Aggregate all the lesion features for a patient using a mean aggregation and process the aggregated feature by traditional machine learning algorithms such as random forest (RF), support vector machine (SVM) with the RBF-kernel, and logistic regression (LR). Table 2 compares our model's performance against these alternatives. Our proposed solution obtains better AUC than the alternatives, indicating that incorporating the graph structure is beneficial for eventual prediction.

Table 2. Effectiveness of the graph structure. We compare our method to traditional ML algorithms and a set-based aggregation baseline, both of which discard the graph structure. The incorporation of the graph structure is beneficial for downstream prediction.

Methods	One year	Two year
	AUC	AUC
SVM	0.635 ±0.05	0.513 ±0.09
RF	0.639 ±0.04	0.650 ±0.09
LR	0.658 ±0.05	0.655 ±0.07
Set-Proc	0.654 ±0.06	0.658 ±0.08
Ours	**0.671 ±0.06**	**0.664 ±0.06**

Table 3. Importance of spatial proximity. GAT (partially) and TransformerConv (completely) ignore spatial proximity in the graph while GCN, EdgeConv, and GraphSAGE incorporate it. The incorporation of spatial proximity is beneficial for downstream prediction.

Methods	One year	Two year
	AUC	AUC
Tra. Conv [18]	0.632 ± 0.06	0.627 ± 0.06
GAT	0.624 ±0.08	0.631 ±0.09
Edge	0.640 ±0.09	0.657 ±0.09
SAGE	0.650 ±0.05	0.634 ±0.06
GCN(Ours)	**0.671 ±0.06**	**0.664 ±0.06**

Table 4. Comparison of the performance of different message passing layers with and without the self-pruning module. Incorporating the self-pruning module is beneficial for most message passing layers.

Methods	One year		Two year	
	w/o SPM	w/ SPM	w/o SPM	w/ SPM
GAT	**0.628 ±0.07**	0.624 ±0.08	0.607 ±0.06	**0.631±0.09**
Edge	0.629 ±0.07	**0.640 ±0.09**	0.651 ±0.07	**0.657 ±0.09**
Sage	0.647 ±0.06	**0.650 ±0.05**	0.630 ±0.07	**0.634 ±0.06**
GCN(Ours)	0.640 ±0.04	**0.671 ±0.06**	0.643 ± 0.07	**0.664±0.06**

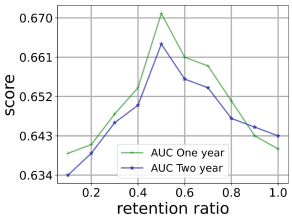

Fig. 2. Model performance against different retention ratio r. *Best performance observed for* $r = 0.5$.

The Importance of Encoding Spatial Proximity. The spatial proximity in our model is encoded at two levels. First, lesion connectivity is determined using a NN graph, and second, we weigh the edges based on their distance. Graph convolution layers such as EdgeConv, GCN, and GraphSAGE take the edge weights into account. (EdgeConv does it implicitly by taking a difference of lesion features that already contain spatial information).

On the other hand, the GAT model learns an attention weight and ignores the pre-defined edge weights. However, it still computes these coefficients on only the connected nodes. We can go further, completely ignore the distances and instead use a fully connected graph. The TransformerConv [18] on such a graph is equivalent to applying the well-known transformer encoder [22] on the inputs. Table 3 shows that the methods that ignore spatial proximity (TransformerConv) or do not use distance-based weighting (GAT) struggle. On the other hand, EdgeConv, GCN, and GraphSAGE work better. We use GCN in our model owing to its superior performance.

The Contribution of the Self-Pruning Module (SPM). The SPM selects a subset of lesions for the final prediction during the training and evaluation phases. However, the proposed classification method can work without it. In this case, none of the lesions is discarded during the readout operation. Table 4 shows the classification results with and without the SPM. We observe that including SPM leads to better outcomes across different message-passing networks. An explanation could be that the SPM can better handle patient variations (in terms of the total number of lesions) by operating on a subset of lesions (Fig. 3).

Analysis of Hyperparameters. The retention ratio r and the number of neighbors k used for building the graph are the two critical hyperparameters in our proposed framework. We discuss the effect of the retention ratio r here and defer discussion about k to the appendix.

Effect of Retention Ratio r. The retention ratio $r \in (0, 1]$ controls the fraction of lesions retained after the self-pruning module. If we set its value to 1, all the lesions are retained for the final prediction and thus, bypassing the self-pruning

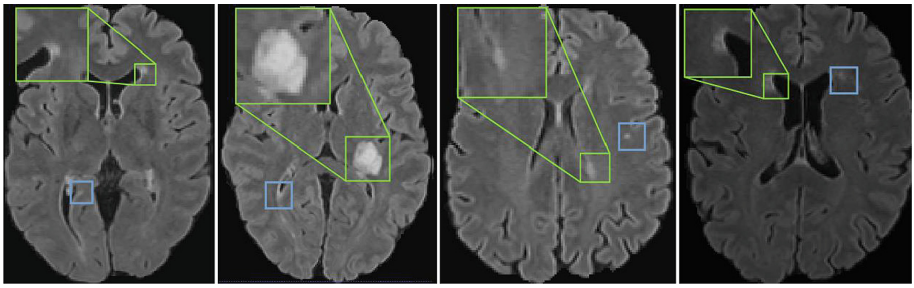

Fig. 3. Lesions selected by the *SPM* for two-year inflammatory disease activity prediction are highlighted with a green bounding box. We also show the zoomed-in view of the lesion. A concurrent lesion in the scan ignored by the *SPM* is shown with a blue bounding box. According to their size and location, the selected lesions are most likely to be relevant to the prediction. (Color figure online)

module. Any other value implies that we ignore at least a few lesions in the readout layer. Since the number of lesions can vary across graphs, we retain $\lceil (N.r) \rceil$ lesions after the self-pruning layer. To find the optimal r, we test our model with r between 0.1 and 1.0. The results are summarized in Fig. 2. We set r to 0.5 for both tasks.

5 Conclusion

Predicting MS inflammatory disease activity is a clinically relevant, albeit challenging task. In this work, we propose a two-stage graph-based pipeline that surpasses existing CNN-based methods by decoupling the tasks of detecting and learning rich semantic features for lesions. We also propose a self-pruning module that further improves model generalizability by handling variations in the number of lesions within patients. Most importantly, we frame the MS inflammatory disease activity prediction as a graph classification problem. We hope our work provides a new perspective and leads to cutting-edge research at the intersection of graph processing and MS inflammatory disease activity prediction.

Acknowledgement. This work was supported by Helmut Horten Foundation. B.W., M.M., and B.M. were supported through the DFG, SPP Radiomics. H.B.L. is supported by an Nvidia GPU research grant.

References

1. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? arXiv preprint arXiv:2105.14491 (2021)
2. 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)

3. Durso-Finley, J., Falet, J.P., Nichyporuk, B., Douglas, A., Arbel, T.: Personalized prediction of future lesion activity and treatment effect in multiple sclerosis from baseline MRI. In: International Conference on Medical Imaging with Deep Learning, pp. 387–406. PMLR (2022)
4. Falet, J.P.R., et al.: Estimating individual treatment effect on disability progression in multiple sclerosis using deep learning. Nat. Commun. **13**(1), 5645 (2022)
5. Filippi, M., et al.: Identifying progression in multiple sclerosis: new perspectives. Ann. Neurol. **88**(3), 438–452 (2020)
6. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
7. Hapfelmeier, A., et al.: Retrospective cohort study to devise a treatment decision score predicting adverse 24-month radiological activity in early multiple sclerosis. Ther. Adv. Neurol. Disord. **16**, 17562864231161892 (2023)
8. Hauser, S.L., Cree, B.A.: Treatment of multiple sclerosis: a review. Am. J. Med. **133**(12), 1380–1390 (2020)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods **18**(2), 203–211 (2021)
10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
11. Li, H., et al.: Imbalance-aware self-supervised learning for 3D radiomic representations. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12902, pp. 36–46. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87196-3_4
12. Liu, C.M., Ta, V.D., Le, N.Q.K., Tadesse, D.A., Shi, C.: Deep neural network framework based on word embedding for protein glutarylation sites prediction. Life **12**(8), 1213 (2022)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Prabhakar, C., Li, H.B., Yang, J., Shit, S., Wiestler, B., Menze, B.: ViT-AE++: improving vision transformer autoencoder for self-supervised medical image representations. arXiv preprint arXiv:2301.07382 (2023)
15. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 652–660 (2017)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Schell, M., et al.: Automated brain extraction of multi-sequence MRI using artificial neural networks. European Congress of Radiology-ECR 2019 (2019)
18. Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., Sun, Y.: Masked label prediction: unified message passing model for semi-supervised classification. arXiv preprint arXiv:2009.03509 (2020)
19. Sormani, M.P., Bruzzi, P.: MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. Lancet Neurol. **12**(7), 669–676 (2013)
20. Sormani, M.P., De Stefano, N.: Defining and scoring response to IFN-β in multiple sclerosis. Nat. Rev. Neurol. **9**(9), 504–512 (2013)
21. Tousignant, A., Lemaître, P., Precup, D., Arnold, D.L., Arbel, T.: Prediction of disease progression in multiple sclerosis patients using deep learning analysis of MRI data. In: International Conference on Medical Imaging with Deep Learning, pp. 483–492. PMLR (2019)

22. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
23. Wang, H., et al.: Mixed transformer U-Net for medical image segmentation. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2390–2394. IEEE (2022)
24. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. (TOG) **38**(5), 1–12 (2019)
25. Wattjes, M.P., et al.: 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. Lancet Neurol. **20**(8), 653–670 (2021)
26. Yoo, Y., et al.: Deep learning of brain lesion patterns for predicting future disease activity in patients with early symptoms of multiple sclerosis. In: Carneiro, G., et al. (eds.) LABELS/DLMIA -2016. LNCS, vol. 10008, pp. 86–94. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46976-8_10
27. Zhang, X., He, L., Chen, K., Luo, Y., Zhou, J., Wang, F.: Multi-view graph convolutional network and its applications on neuroimage analysis for Parkinson's disease. In: AMIA Annual Symposium Proceedings, vol. 2018, p. 1147. American Medical Informatics Association (2018)
28. Zhang, Y.D., Pan, C., Sun, J., Tang, C.: Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. J. Comput. Sci. **28**, 1–10 (2018)