Multi-modal Semi-supervised Evidential Recycle Framework for Alzheimer's Disease Classification

Yingjie Feng[1], Wei Chen[2], Xianfeng Gu[3], Xiaoyin Xu[4], and Min Zhang[1(✉)]

[1] Collaborative Innovation Center of Artificial Intelligence, College of Computer Science and Technology, Zhejiang University, Hangzhou, China
min_zhang@zju.edu.cn
[2] Zhejiang University Affiliated Sir Run Run Shaw Hospital, Hangzhou, China
[3] Department of Computer Science, Stony Brook University, Stony Brook, NY, USA
[4] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract. Alzheimer's disease (AD) is an irreversible neurodegenerative disease, so early identification of Alzheimer's disease and its early stage disorder, mild cognitive impairment (MCI), is of great significance. However, currently available labeled datasets are still small, so the development of semi-supervised classification algorithms will be beneficial for clinical applications. We propose a novel uncertainty-aware semi-supervised learning framework based on the improved evidential regression. Our framework uses the aleatoric uncertainty (AU) from the data itself and the epistemic uncertainty (EU) from the model to optimize the evidential classifier and feature extractor step by step to achieve the best performance close to supervised learning with small labeled data counts. We conducted various experiments on the ADNI-2 dataset, demonstrating the effectiveness and advancement of our method.

Keywords: Semi-supervised learning · Deep evidential regression · EfficientNet-V2 · Alzheimer's disease · Multi-modality

1 Introduction

Alzheimer's disease (AD) is an irreversible neurodegenerative disease that leaves patients with impairments in memory, language and cognition [7]. Previous work of [6,22] show that the combination of image data and other related data is beneficial to the improvement of model performance, but how to efficiently combine statistical non-imaging data and medical image data is still an open question. Second, although it is not too difficult to obtain and collect patient data, subjective bias in the AD diagnosis process and the time-consuming and complicated

process of labeling the diagnostic results lead to the scarcity of labeled data [11]. Therefore, research and development of models that require only a small amount of labeled data to achieve higher accuracy has attracted great attention [14].

Semi-supervised learning (SSL) methods are commonly used in medical image analysis to address the lack of manually annotated data [24]. Hang et al. [10] proposed a contrastive self-ensembling framework by introducing the weight formula and reliability-awareness for semi-supervised medical image classification. In [3], Aviles et al., based on the diffusion model and hypergraph learning, proposed a multi-modal hypergraph diffusion network to implement semi-supervised learning for AD classification. In [4], researchers introduced CSEAL, a semi-supervised learning framework that combines consistency-based SSL with uncertainty-based active learning, for multi-label chest X-ray classification tasks. Other studies using evidential learning [5,17] have demonstrated the great potential of this theory in fitting low-dimensional manifolds in high-dimensional spaces for classification with uncertainty estimates. This feature makes the model based on evidential learning promising in the SSL field of medical images.

The proposal of evidential deep learning (EDL) [21] allows the model to better estimate the uncertainty in multi-classification tasks. On the binary classification task, controlling evidential regression to obtain a continuous probability value before 0 and 1 can often achieve more accurate results than using the Dirichlet distribution to obtain a discrete distribution of EDL [19].
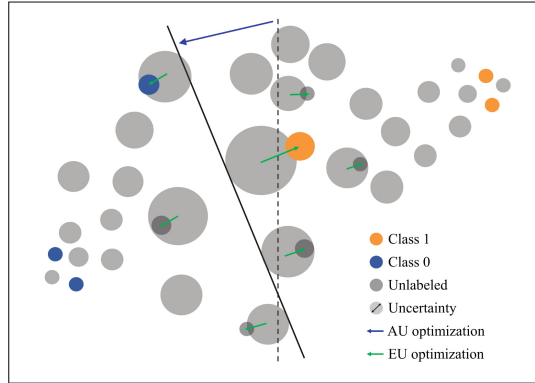

Fig. 1. An iteration of optimization of our model. The AU optimization process does not rely on any label and obtains the best classification of the current model by reducing AU. Then, on the basis of this classification, EU optimization relies on ground truth and pseudo labels to optimize the model to get better prediction results.

The residual between the prediction results of the imperfect model and the true distribution of the data can be decomposed into aleatoric uncertainty (AU) and epistemic uncertainty (EU). Theoretically, the former comes from the noise of the data, which usually does not depend on the sample size. Therefore, by

iteratively reducing this part of uncertainty, the best classification results can be obtained under a given amount of data. The latter is proportional to the sample size. As sample size increases, the reduction of this part of uncertainty can make the model closer to the observed distribution or fit with more complex conditions, thereby improving the performance of the model itself. Based on this understanding, we exploit the ability of evidential regression of handling uncertainty to decompose the two parts of uncertainty, AU and EU, and proposed a method by adjusting the two parts of uncertainty to achieve semi-supervised classification which shows in Fig. 1.

Our main contributions include: 1) Adjusting the loss function of evidential regression so it can obtain more accurate results and better separate AU and EU; 2) Building a multi-layer and multi-step network to implement evidential regression and a semi-supervised learning method of step-by-step training is proposed; 3) A new SOTA of semi-supervised learning is achieved on the ADNI dataset, and performance close to supervised learning can be achieved with only a small amount of labeled data.

2   Methods

2.1   Original Deep Evidential Regression (DER)

DER [2] adopts the simplest setting: $y_i \sim N(0, \sigma_i^2)$. In a Bayesian framework, this corresponds to taking the normal inverse Gamma distribution $\mathrm{NIG}(\mu, \sigma^2|m)$, $m = (\gamma, \nu, \alpha, \beta)$, as a conjugate prior of a normal contribution with unknown mean μ and variance σ^2. Combining the disturbance parameter with Bayesian inference, the likelihood of an observation y for a given m follows a t-distribution with $2\alpha_i$ degrees of freedom: $L_i^{NIG} = St_{2\alpha_i}\left(y_i|\gamma_i, \frac{\beta_i(1+\gamma_i)}{\gamma_i \alpha_i}\right)$. For known m, Animi et al. [2] defined the prediction of y_i as $E[\mu_i] = \gamma_i$, and defined AU and EU as u_a and u_e:

$$u_a^2 = E[\sigma_i^2] = \beta_i/(\alpha_i - 1), \qquad u_e^2 = \mathrm{var}[\mu_i] = E[\sigma_i^2]/\nu_i.$$

And it follows that:

$$L_i^{'}(w) = -\log L_i^{NIG}(w) + \lambda L_i^R(w), \quad L_i^R(w) = |y_i - \gamma_i| \cdot \Phi$$

where $m = NN(w)$ is specified by a neural network (NN), λ is a hyperparameter, and $\Phi = 2\gamma_i + \alpha_i$ represents the total evidence gained from training.

2.2   Evidential Regression Beyond DER

Although DER has achieved some success in both theoretical and practical applications [2,15], as pointed out by Meinert et al. [16], this theory has some major flaws. First, although the regularization part of loss function L^R is added, the constrain on parameter β_i is not enough. Second, although the two parts of AU and EU are defined separately, the correlation between them is too high.
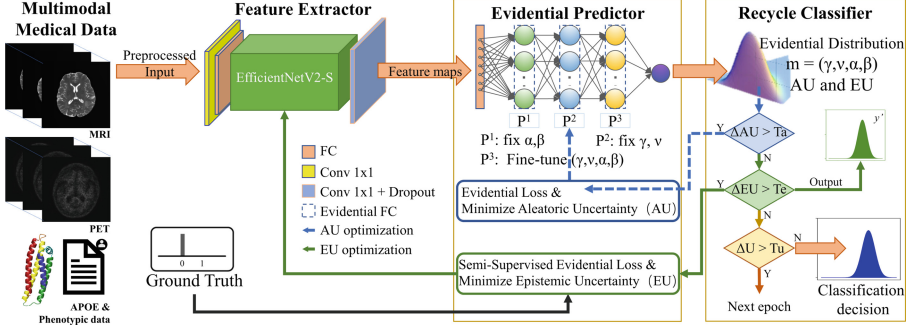
Fig. 2. Architecture of our network. The recycle classifier part judges the decrease of the uncertainty of the predicted value, and controls the framework to train the other two parts in a loop.

In practice, disentangling and effectively using uncertainty information for training remains challenging.

After practice and theoretical proof, Meinert et al. [16] states that the width of the t-distribution projected by the NIG distribution, that is, w_{St}, can better reflect the noise in data.

And, correspondingly, we use the residual $1/\sqrt{\nu_i}$ part of u_a and u_e in the original definition to represent EU:

$$u_A = w_{St} = \sqrt{\frac{\beta_i(1+\nu_i)}{\alpha_i\nu_i}}, \qquad u_E = \frac{u_e}{u_a} = \frac{1}{\sqrt{\nu_i}}$$

where ν_i, α_i, and β_i are part of the parameters of the evidence distribution $m = (\gamma, \nu, \alpha, \beta)$, and we verify the performance of this new uncertainty estimation method through experiments.

2.3 Model and Workflow

With efficient estimation of AU and EU, our model has the basis for implementation. As shown in Fig. 2, our model is divided into three parts: multimodal feature extractor, evidential predictor, and recycle classifier. The multimodal feature extractor form a high-dimensional feature space and the evidential predictor generates the evidential distribution in the feature space. After calculating the classification result and the uncertainty (AU and EU) based on the evidential distribution, the recycle classifier controls the training process and reaches the best performances through a step-by-step recurrent training workflow.

AU for Training Classifier. Based on the manifold assumption, the real data is gathered on the low-dimensional manifold of the high-dimensional space, and the noise of the data is located on the edge of the manifold for the corresponding

category. When using ER to fit the manifold, these noise data will make marginal data with high AU. By optimizing the classifier to iteratively reduce the AU, optimal classification result under the current conditions can be obtained.

We use L^a to optimize AU:

$$L_i^a(w) = -\log L_i^{NIG'}(w) + \lambda_a \left(\frac{\gamma_i - \gamma_i^2}{w_{St}^2} \right) \Phi'.$$

In the above formula $\lambda_a = [0.005, 0.01]$ is a parameter that controls the degree of deviations of the regularization part and w_{St} uses the previous definition, and Φ' is the total amount of evidence learned by the model. In order to better motivate the learning of the model, we adopted the work of Liu et al. [15] and used the form of $\Phi' = \gamma_i + 2\alpha_i$. We used the expanded form of L^{NIG} with minor adjustments according to the optimization objective, specifically, $-\log L_i^{NIG'} = \frac{1}{2}\log(\frac{\pi}{\nu}) - \alpha\log(2\beta + 2\beta\nu) + (\alpha + 0.5)\log(\nu(\gamma_i - \gamma_i^2) + 2\beta + 2\beta\nu) + \log\left(\frac{\Gamma(\alpha)}{\Gamma(\alpha+0.5)}\right)$.

EU for Training Extractor. If only the AU part is optimized, there will always be this gap between the model prediction and the real data.

EU is mainly used to optimize the feature extractor since EU mainly reflects the bias of the model in the prediction. For data D_l, given groundtruth labels, we use $L_i^l(w) = -\log L_i^{NIG}(w) + \lambda_l \left(\frac{y_i - \gamma_i}{w_{St}}\right)^2 \Phi'$. In order to enhance the certainty of the data with ground truth, we set a smaller $\lambda_l = [0.005, 0.015]$. For dataset D_u without real labels, we use the prediction results y' obtained in the last iterative training to replace the real labels to get $L_i^u(w) = -\log L_i^{NIG}(w) + \lambda_u \left(\frac{y_i' - \gamma_i}{w_{St}}\right)^2 \Phi'$. In order for the model to utilize the results of the previous round of learning, we set a larger $\lambda_u = [0.015, 0.025]$, which can make the model more conservative about making predictions in the next iteration. This reduces our models being affected by misleading evidence and obtains better performance by retaining higher uncertainty to allow the model to have more room to optimize.

In order to effectively combine labeled and unlabeled data we adjust the weights of different data:

$$L_i^e(w) = \mu_l L_i^l + \mu_u L_i^u = -\log L_i^{NIG}(w) + \mu_u \lambda_u \left(\frac{y_i' - \gamma_i}{w_{St}}\right)^2 \Phi' + \mu_l \lambda_l \left(\frac{y_i - \gamma_i}{w_{St}}\right)^2 \Phi'$$

where $\mu_l + \mu_u = 1$, $\mu_l, \mu_u \in [0, 1]$, are two weight factors.

Model. In terms of the feature extractor, we use the latest EfficientNetV2, which, in Feng et al. [8], has achieved good results in combination with EDL. In order to avoid overfitting, we used the minimum model in this network and added Dropout to the output end. At the same time, in order to fill the differences between multi-modality data and model input, we have added the fully connected (FC) layer and convolutional layer (Conv) to adaptive adjust input channels. We employed three evidential FC layers proposed by Amini et al. [2] to form our

evidential predictor. At the same time, in order to achieve the optimization of AU and EU separately, we froze some parameters in the first two layers and limited the range of the last layer of parameter adjustment.

Workflow of Recycle Training. First, we do not fix any weight and we use a small part of the data for warm-up training. Second, we freeze the weight update of the extractor and P^2 and P^3 in the evidential predictor and use L^a to optimize the classifier. We calculate and record the AU score after each update. When the difference between the update $|\Delta AU|$ is smaller than the threshold value $T_a = [0.0005, 0.001]$ we set, the cycle of AU optimization is over. Then, we fix the weight of P^1 and P^3 in the evidential predictor and use L^e to optimize the extractor. Similarly, when the change $|\Delta EU|$ brought by the update is less than the threshold $T_e = [0.0025, 0.005]$, end the cycle and output y'. Finally, we fix all network parameters except P^3 to fine-tune until $|\Delta U| = |\Delta AU| + |\Delta EU|$ brought by the update loss function $L = L^e + L^a$ is less than threshold $T_u = [0.002, 0.005]$. All thresholds are adjusted according to the proportion of labels and unlabeled data during training.

Table 1. Classification results of all comparison methods on ADNI-2 dataset (%). The top section of the table shows the results of supervised learning (SL), while the bottom section shows the performance of the current SSL SOTA methods.

Method	AD vs. NC			EMCI vs. LMCI			LMCI vs. NC		
	ACC	SPE	SEN	ACC	SPE	SEN	ACC	SPE	SEN
Baseline [12]	80.53	80.10	80.32	74.10	73.18	75.85	72.05	70.80	71.26
SL SOTA [18,20,23]	96.84	98.23	95.76	92.40	93.70	89.50	92.49	91.08	93.48
Upper bound (UB)	94.45	93.80	94.07	89.95	90.29	90.81	88.56	88.81	86.32
Π model [13]	90.45	85.82	90.05	81.58	80.15	84.50	80.65	83.48	78.75
DS^3L [9]	90.86	89.03	89.72	81.07	83.25	82.81	80.79	81.55	81.16
RFS-LDA [1]	92.11	89.50	88.40	80.90	81.05	83.63	81.90	84.72	80.05
Hypergraph diffusion [3]	92.11	92.80	91.33	85.22	86.40	84.02	82.01	84.01	81.80
Ours	**93.90**	**92.95**	**93.01**	**89.45**	**88.50**	**89.47**	**87.27**	**86.94**	**85.83**

3   Experiments and Results

Data Description. In this paper, we assess the effectiveness of our multi-modal semi-supervised evidential recycle framework on the ADNI-2 dataset[1], which comprises multi-center data consisting of various modalities, including imaging and multiple phenotype data. Specifically, the dataset consists of four

---

[1] *Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

categories: normal control (NC), early mild cognitive impairment (EMCI), late mild cognitive impairment (LMCI), and Alzheimer's disease (AD). To ensure the effectiveness of our training and balance the number of categories, we used a sample of 515 patients, utilizing their MRI, PET, demographics, and APOE as inputs. On MRI images, we used 3T T1-weighted and FLAIR MR images, and the preprocessing process used CAT12 and SPM tools. All MRI data were processed using standard pipeline, including anterior commissure (AC)-posterior commissure (PC) correction, intensity correction, and skull stripping. Affine registration is performed to linearly align each MRI to the Colin27 template and resample to $224 \times 224 \times 91$ for subsequent processing. For PET images, we used the official pre-processed AV-45 PET image and resampled them in the same way as the MRIs. We chose to include APOE in our analysis, as it is a well-established genetic risk factor for developing AD.

Evaluation. We evaluated our model from three aspects. First, for the sake of comparison, we followed the technical conventions of most similar studies and selected three comparison tasks: AD vs NC, LMCI vs NC, and EMCI vs LMCI. Second, we compared and demonstrated the results of our model under different numbers of ground truth labels to verify that its performance improves as the label data volume increases. Third, we conducted different ablation experiments, which shows in Fig. 3, to prove the validity and rationality of each part of the proposed model framework. Among them, CNNs represents the performance when using only the EfficientNetV2-S model and its original supervised learning classifier without using unlabeled data for training, which is the baseline model. AU and EU represent the training process using only the corresponding parts. DER uses our proposed complete training process but does not use our improved u_a and u_e estimations, instead continuing to use the estimation method u_A and u_E which proposed in the original DER paper [2]. To compare performance fairly, we ran all techniques under the same conditions. The results were evaluated on accuracy (ACC), specificity (SPE), and sensitivity (SEN).

Implementation Details. The upper bound in performance is the result obtained when the model is trained with all the input data are labeled. In the current supervised learning algorithms, the performance of each algorithm on
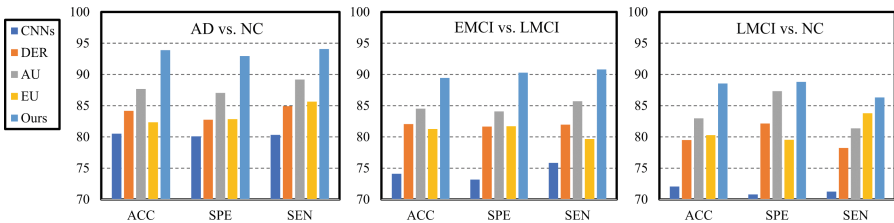

Fig. 3. Ablation experiment results.

Table 2. Classification results(%) with different percentages of labeled and unlabelled data in the training process.

Method	Labeled	AD vs. NC			EMCI vs. LMCI			LMCI vs. NC		
		ACC	SPE	SEN	ACC	SPE	SEN	ACC	SPE	SEN
Upper Bound	100%	94.45	93.80	94.07	89.95	90.29	90.81	88.56	88.81	86.32
Baseline	5%	75.11	73.92	75.79	73.17	69.47	75.14	68.95	67.81	68.53
DS^3L		78.92	76.75	78.14	74.68	71.65	72.39	74.27	70.58	73.71
Ours		82.45	81.63	84.97	73.05	70.84	74.22	73.86	74.15	72.28
Baseline	10%	78.50	78.03	77.48	73.67	70.29	74.11	69.13	68.57	68.38
DS^3L		84.73	79.67	82.69	80.93	80.13	79.54	74.57	75.39	72.88
Ours		90.18	89.73	87.42	81.67	83.45	82.29	80.01	78.25	80.89
Baseline	20%	80.53	80.10	80.32	74.10	73.18	75.85	72.05	70.80	71.26
DS^3L		90.86	89.03	89.72	81.07	83.25	82.81	80.79	81.55	81.16
Ours		**93.90**	**92.95**	**93.01**	**89.45**	**88.50**	**89.47**	**87.27**	**86.94**	**85.83**

each task is not consistent, so we selected three papers in supervised learning, each representing the SOTA performance of the three tasks [18, 20, 23] for comparison. Our implementation employs PyTorch v1.4.0 and utilizes the Adam optimizer with a learning rate of 1×10^{-4} and a weight decay of 1×10^{-5}. We utilize a linear decay scheduler of 0.1 based on the loss functions above. The optimizer is set with β values of [0.9, 0.999] and ϵ value of 1×10^{-8}. In terms of data, since the SSL method needs to learn from unlabeled data, 100% of the data is put into training, and some of the data have ground truth labels. In the test, only the result index of the unlabeled data is calculated, so the training set and the test set are not divided. But in order to determine the threshold of each uncertainty, we randomly selected 10% of the data as the validation set, and calculated the uncertainty independently outside the training process.

Results. We compared our model with the semi-supervised learning methods currently achieving the best performance on the ADNI-2 dataset, as well as other top models in the semi-supervised learning field. As shown in Table 1, our model achieved SOTA performance in all three tasks of the semi-supervised learning category. At the same time, compared with other semi-supervised learning algorithms, our results are unprecedentedly close to the best supervised learning methods, indicating the performance of our model under less labeled data and the feasibility of applying this algorithm in clinical settings.

Our ablation experiment results are shown in Fig. 3. Firstly, compared with the baseline, our semi-supervised learning algorithm effectively learns classification information from unlabeled data. Secondly, compared with DER, our uncertainty estimation surpasses the original DER method. The AU and EU items demonstrate the importance of optimizing both the AU and EU components in our framework.
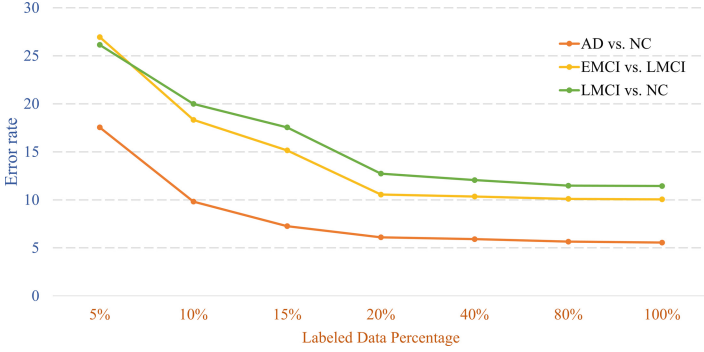
Fig. 4. Error rate of different percentages of label counts, the well-known conduction effect in the field of semi-supervised learning can be observed.

From Table 2, we can observe that we have outperformed the currently representative advanced semi-supervised learning algorithm DS^3L [9] in each labeled data count. At the same time, the superiority of our model compared to the baseline method also proves the learning efficiency of our framework. The performance of our model at 20% labeled data count is already very close to the upper bound, which is the result obtained using 100% labeled data. This indicates the strong learning ability of our model in the case of a small labeled data amount.

In addition, we have plotted the error rate of our framework under different labeled data counts in Fig. 4. It is apparent that the performance of our model improves as the labeled data amount increases from 5% to 10%, 15%, and 20%. Combining with Table 2, we can observe the well-known transductive effect in the field of semi-supervised learning, which means that beyond a certain data amount, increasing the size of the dataset can only bring marginal performance improvement. This is evident when comparing the model performance under 20%, 40%, 80%, and 100% labeled data counts.

4 Conclusions

We proposed an evidential regression-based semi-supervised learning framework, using the characteristics of AU and EU to train classifiers and extractors, respectively. Our model achieves SOTA performance on the ADNI-2 dataset. And due to the characteristics of semi-supervised learning, our model has unique advantages in adding private data, fine-tuning downstream tasks, and avoiding overfitting, which makes our model have great potential in clinical applications.

References

1. Adeli, E., et al.: Semi-supervised discriminative classification robust to sample-outliers and feature-noises. IEEE Trans. Pattern Anal. Mach. Intell. **41**(2), 515–522 (2018)

2. Amini, A., Schwarting, W., Soleimany, A., Rus, D.: Deep evidential regression. Adv. Neural. Inf. Process. Syst. **33**, 14927–14937 (2020)

3. Aviles-Rivero, A.I., Runkel, C., Papadakis, N., Kourtzi, Z., Schönlieb, C.B.: Multi-modal hypergraph diffusion network with dual prior for Alzheimer classification. In: MICCAI 2022, Part III. LNCS, pp. 717–727. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_69

4. Balaram, S., Nguyen, C.M., Kassim, A., Krishnaswamy, P.: Consistency-based semi-supervised evidential active learning for diagnostic radiograph classification. In: MICCAI 2022, Part I. LNCS, vol. 13431, pp. 675–685. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16431-6_64

5. Bengs, V., Hüllermeier, E., Waegeman, W.: Pitfalls of epistemic uncertainty quantification through loss minimisation. In: Advances in Neural Information Processing Systems (2022)

6. Cobbinah, B.M., et al.: Reducing variations in multi-center Alzheimer's disease classification with convolutional adversarial autoencoder. Med. Image Anal. **82**, 102585 (2022)

7. De Strooper, B., Karran, E.: The cellular phase of Alzheimer's disease. Cell **164**(4), 603–615 (2016)

8. Feng, Y., Wang, J., An, D., Gu, X., Xu, X., Zhang, M.: End-to-end evidential-efficient net for radiomics analysis of brain MRI to predict oncogene expression and overall survival. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, Part III, vol. 13433, pp. 282–291. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_27

9. Guo, L.Z., Zhang, Z.Y., Jiang, Y., Li, Y.F., Zhou, Z.H.: Safe deep semi-supervised learning for unseen-class unlabeled data. In: International Conference on Machine Learning, pp. 3897–3906. PMLR (2020)

10. Hang, W., Huang, Y., Liang, S., Lei, B., Choi, K.S., Qin, J.: Reliability-aware contrastive self-ensembling for semi-supervised medical image classification. In: Medical Image Computing and Computer Assisted Intervention-MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part I. pp. 754–763. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16431-6_71

11. Hett, K., Ta, V.T., Oguz, I., Manjón, J.V., Coupé, P., Initiative, A.D.N., et al.: Multi-scale graph-based grading for Alzheimer's disease prediction. Med. Image Anal. **67**, 101850 (2021)

12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Cision and Pattern Recognition, pp. 4700–4708 (2017)

13. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. In: International Conference on Learning Representations (2016)

14. Li, Z., Togo, R., Ogawa, T., Haseyama, M.: Chronic gastritis classification using gastric x-ray images with a semi-supervised learning method based on tri-training. Med. Biol. Eng. Comput. **58**, 1239–1250 (2020)

15. Liu, Z., Amini, A., Zhu, S., Karaman, S., Han, S., Rus, D.L.: Efficient and robust lidar-based end-to-end navigation. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 13247–13254. IEEE (2021)

16. Meinert, N., Gawlikowski, J., Lavin, A.: The unreasonable effectiveness of deep evidential regression. arXiv e-prints pp. arXiv-2205 (2022)
17. Neupane, K.P., Zheng, E., Yu, Q.: MetaEDL: Meta evidential learning for uncertainty-aware cold-start recommendations. In: 2021 IEEE International Conference on Data Mining (ICDM), pp. 1258–1263. IEEE (2021)
18. Ning, Z., Xiao, Q., Feng, Q., Chen, W., Zhang, Y.: Relation-induced multi-modal shared representation learning for Alzheimer's disease diagnosis. IEEE Trans. Med. Imaging **40**(6), 1632–1645 (2021)
19. Oh, D., Shin, B.: Improving evidential deep learning via multi-task learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, pp. 7895–7903 (2022)
20. Pei, Z., Wan, Z., Zhang, Y., Wang, M., Leng, C., Yang, Y.H.: Multi-scale attention-based pseudo-3D convolution neural network for Alzheimer's disease diagnosis using structural MRI. Pattern Recogn. **131**, 108825 (2022)
21. Sensoy, M., Kaplan, L., Kandemir, M.: Evidential deep learning to quantify classification uncertainty. In: Advances in Neural Information Processing Systems 31 (2018)
22. Song, X., et al.: Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction. Med. Image Anal. **69**, 101947 (2021)
23. Song, X., et al.: Multi-center and multi-channel pooling GCN for early AD diagnosis based on dual-modality fused brain network. IEEE Trans. Med. Imaging (2022)
24. Yang, X., Song, Z., King, I., Xu, Z.: A survey on deep semi-supervised learning. IEEE Trans. Knowl. Data Eng. (2022)