Neuro-AMS: Neuro-informed Age-aware and Medical Knowledge-integrated Strategy for Diagnosis of Multiple Brain Disorders

Zhenguo Zhang[1], Lin Teng[1], Nan Zhao[1], Yuxiao Liu[1], Zhaoyu Qiu[1], Zehao Weng[1], Jinwei Kong[1], Feng Shi[2(✉)], Dinggang Shen[1,2,3(✉)]

[1]School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
[2]Department of Research and Development, United Imaging Intelligence, Shanghai 200230, China
feng.shi@uii-ai.com
[3]Shanghai Clinical Research and Trial Center, Shanghai 201210, China
dgshen@shanghaitech.edu.cn

Abstract. Neurological diseases encompass a diverse range of conditions such as neurodegenerative diseases and neurodevelopmental disorders. Developing a general model to assist in the diagnosis of multiple neurological diseases is essential in clinical practice, as it can help reduce misdiagnosis rates and alleviate the burden on physicians. However, most existing diagnostic models are designed for specific neurological disease scenarios and show poor performance when applied to multiple diseases. To this end, we present a semantic-assisted framework, called **Neuro-AMS**, a **Neuro**-informed **A**ge-aware and **M**edical knowledge-integrated **S**trategy for diagnosis of multiple brain disorders. Specifically, we employ a vision encoder based on age-aware strategy to further enhance performance by leveraging the potential relationship between age and neurological diseases. Additionally, we extract semantic features from labels and integrate corresponding medical knowledge embeddings, constructing knowledge-level label features with enhanced semantics. These knowledge-level label features guide the vision encoder for capturing higher-level semantic representations through the alignment of image-text pairs. Our method is evaluated on four public brain disease datasets, and experimental results demonstrate that our method achieves consistent and statistically significant improvement compared with three public benchmarks and three specialized models.

Keywords: Neurological disease diagnosis · Vision language models · Medical domain knowledge · Age-aware strategy.

1 Introduction

Neurological diseases involve a variety of conditions, including neurodegenerative diseases and neurodevelopmental disorders. These conditions bring a devastating

effect on the central nervous system, contributing significantly to poor health and disability worldwide. According to a study by Steinmetz *et al.* [19] over 3 billion people worldwide were affected by nervous system disorders in 2021. Notably, neurological diseases such as Alzheimer's Disease (AD) and Autism Spectrum Disorder (ASD) rank among the top ten most prevalent diseases [19]. However, neurological diseases often have a high clinical misdiagnosis rate due to the homogeneous nature of brain Magnetic Resonance Imaging (MRI), which demands expert-level knowledge and diagnostic experience. Therefore, developing an automated diagnostic model for multiple neurological diseases is crucial in clinical practice.

Several deep learning-based methods have emerged and gained wide attention due to their ability to reduce diagnostic costs and improve accuracy [6,12,5]. However, most state-of-the-art (SOTA) deep learning models are designed for specific disease diagnosis tasks [10,13,11], limiting their applicability as comprehensive diagnostic tools for multiple neurological diseases. Furthermore, these models primarily rely on single modality information, overlooking the rich semantic information embedded in disease labels and the valuable medical domain knowledge that could further enhance diagnosis performance.

Recently, Contrastive Language-Image Pre-training (CLIP) [16] has shown great potential, especially in medical image analysis, since it can align image-text information in a shared latent space [20,26]. This capability offers a promising foundation for developing general diagnostic models. While several classification methods leveraging image-text alignment have achieved promising results in natural images and 2D medical images. These Vision-Language Models (VLMs) are predominantly pre-trained on 2D imaging data [16,24,1]. This limitation hinders their ability to handle 3D volumetric data, which is essential for neurological analysis. Furthermore, these methods show poor performance compared to task-specific approaches [22], as they struggle to generate discriminative text representations from limited label information and lack training strategies specifically optimized for classification tasks.

In this paper, we present **Neuro-AMS**, a **Neuro**-informed **A**ge-aware and **M**edical knowledge-integrated **S**trategy for diagnosis of multiple brain disorders. The main idea of this framework is to leverage disease semantics to supervise vision encoder in learning more discriminative representations of disease, while fusing age-guided features as diagnostic priors for neurological diseases. Unlike traditional methods that treat labels as discrete vectors, our approach leverages the semantics of the labels through a pre-trained language model from CLIP. Furthermore, we incorporate disease-specific medical knowledge into label features to construct knowledge-level label representations, which preserves class semantics while leveraging fine-grained domain knowledge to enhance the discrimination of text features. Besides, recognizing the significant correlation between age and neurological diseases [8], we also adopt a vision encoder with age-aware strategy to fully leverage age-related information. Compared to three public benchmarks and three SOTA specialized models, our Neuro-AMS demon-

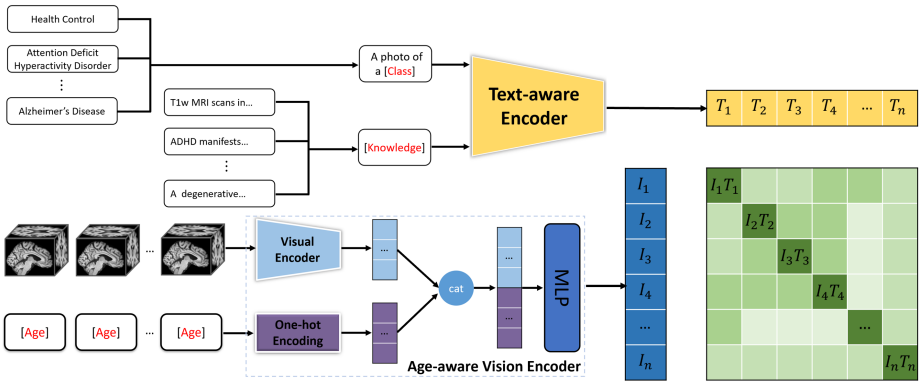

Fig. 1. The overall framework of our proposed Neuro-AMS, consisting of two parts: a Text-aware Encoder and an Age-aware Vision Encoder. Here, we optimize the semantic consistency between knowledge-level label features and age-aware image features, guiding the Age-aware Vision Encoder in capturing more distinguishing features of diseases.

strates consistent and statistically significant improvements across four public datasets with multiple neurological diseases.

2 Methodology

Consider a set of brain MR Image-label pairs (x, y), where $x \in X$ represents a T1-weighted (T1w) MR image, and $y \in Y$ is corresponding disease label from a set of n classes. Our goal is to develop an automatic and robust approach to map each T1w MR image x_i to its corresponding label y_i. The proposed framework shown in Fig. 1 consists of two main components: a Text-aware Encoder and an Age-aware Vision Encoder. The main idea is to utilize meaningful knowledge features extracted by the Text-aware Encoder to guide the age-aware vision encoder for capturing more crucial image features that are highly related to diseases.

2.1 Text-aware Encoder

To effectively utilize text information, we propose a novel Text-aware Encoder to extract text features inspired by TCP [23]. The Text-aware Encoder consists of two pre-trained Text Encoders from CLIP which are responsible for extracting label and knowledge features and a Knowledge Embedding Module (KEM) to integrate domain-specific knowledge (details provided in Fig. 2).

Specifically, the domain knowledge about diseases collected from the Unified Medical Language System (UMLS) [25] is first combined into a sentence (*e.g.* "AD: A degenerative disease of the brain characterized by the insidious onset of

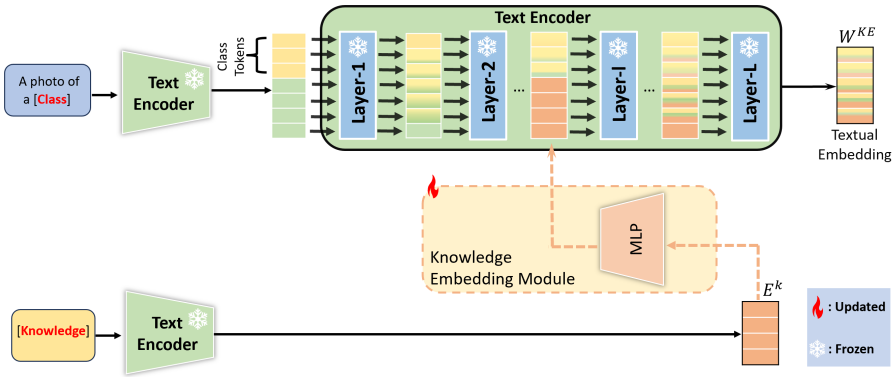

Fig. 2. Architecture of our proposed Text-aware Encoder, which consists of two components: a Text Encoder and a Knowledge Embedding Module (KEM). Knowledge from UMLS is mapped by the KEM to prompt tokens of the corresponding class, generating knowledge-level label embedding W^{KE}.

dementia... The condition primarily occurs after the age of 60."), which is fed into the pre-trained text encoder, to obtain a knowledge embedding $E^k \in \mathbb{R}^{n \times D}$, where n is the number of classes and D is dimension of the text feature. Then, the knowledge embedding E^k is projected into the corresponding class prompt $P_k = F(E^k)$ using our proposed KEM module. In detail, the KEM consists of two hidden layers. In the first layer, the knowledge embedding E^k is projected into a hidden space using a weight matrix $W_1 \in \mathbb{R}^{D \times D_h}$, where D_h is the dimension of the hidden space. In the second layer, the output from the hidden space is mapped to the feature space using a weight matrix $W_2 \in \mathbb{R}^{D_h \times D_{ctx'}}$, where $D_{ctx'}$ is the product of the prompt length L_p and the dimension of the prompt D. Finally, the knowledge embedding $E^k \in \mathbb{R}^{n \times D}$ is projected into knowledge tokens $P_k \in \mathbb{R}^{n \times D_{ctx'}}$, which are then reshaped into $P_k \in \mathbb{R}^{n \times L_p \times D}$ to be inserted into the middle layers of the Text Encoder ϕ.

For the second part, we use a prompt in the form of "A photo of a $[Class]$", where $Class$ corresponds to the disease class. These prompts are encoded by the Text Encoder ϕ. Thus, we can obtain the input textual tokens $T_0 = \{P, C\}$, where $P \in \mathbb{R}^{n \times L_p \times D}$ represents the prompt tokens, and $C \in \mathbb{R}^{n \times 1 \times D}$ is the set of class tokens. These textual tokens T_0 are fed into the first l layers of the Text Encoder ϕ to obtain the intermediate embedding T_l, similar to TCP. Subsequently, the knowledge tokens P_k are inserted into T_l to generate the knowledge-level label tokens $T_l^{'}$,

$$T_l^{'} = [P_{k,1}, P_{k,2}, ..., P_{k,L_p}, T_{l,L_p+1}], \tag{1}$$

where $P_{k,i}$ denotes the $i\text{-}th$ index of P_k in the second dimension, and $T_{l,j}$ denotes the $j\text{-}th$ index of T_l in the second dimension. Finally, the output from the last layer L is treated as the Knowledge-level label Embedding (KE) $W^{KE} \in \mathbb{R}^{n \times D}$.

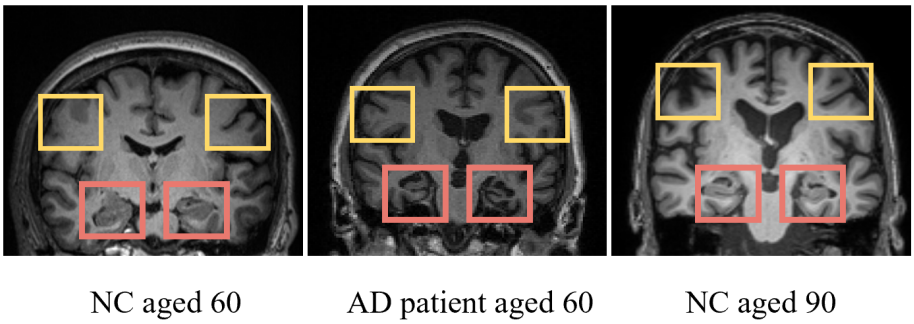

Fig. 3. Comparisons of representative T1w MR images (hippocampi in red boxes, and peripheral brain sulci in yellow boxes) from 60-year-old NC and AD subjects, as well as a 90-year-old NC subject.

2.2 Age-aware Vision Encoder

Most neurological diseases bring prominent changes in local brain structures in T1w MR images, while global changes being relatively less noticeable. These local changes are significant in the early stages of disease, which carry the most diagnostic value [14]. For example, as shown in Fig. 3, the brain structural features of a 60-year-old patient with AD show more pronounced hippocampal atrophy (indicated by the red box) compared to a Normal Control (NC) subject with the same age, while other regions (yellow box) exhibit minimal changes. However, when compared to a 90-year-old NC subject, the differences in the peripheral sulcal regions (yellow box) become more pronounced, while the hippocampal region (red box) shows less noticeable changes due to AD [17].

Based on the above medical knowledge, to build the relationship between brain MR image and age, we propose an Age-aware Vision Encoder based on a novel hierarchical age learning strategy. Specifically, the Age-aware Vision Encoder includes two branches: 1) a pre-trained vision encoder from y-aware contrastive learning [4] that can align age and brain structure in a common latent space and 2) a one-hot encoding module. T1w MR images are fed into the pre-trained vision encoder to extract image features $I_{ori} = \{i_{ori}\}_{j=1}^{n}$, and age information is encoded through one-hot encoding $A = \{a\}_{j=1}^{n}$ based on an age group separation every decade, to obtain high-level age features. These features are concatenated and further fused by an MLP to generate age-aware image features $I = \{i\}_{j=1}^{n}$. Finally, the semantics produced by the Text-aware Encoder is used to supervise the Age-aware vision encoder, further learning the relationships between disease, brain structure, and age-related features.

2.3 Training Strategy

During the training phase, all Age-aware Vision Encoder and Knowledge Embedding Module are updated while freezing the pre-trained Text Encoder from CLIP.

The training is supervised by minimizing the distance between image features I and text features T (i.e., W^{KE}) via cosine similarity, as shown in Eq. (2):

$$s_{i,j} = \frac{I_i \cdot T_j}{\|I_i\| \cdot \|T_j\|} = \frac{\sum_{k=1}^{D} I_i^k \times T_j^k}{\sqrt{\sum_{k=1}^{D}(I_i^k)^2} \times \sqrt{\sum_{k=1}^{D}(T_j^k)^2}}, \tag{2}$$

This produces a similarity matrix $S_{n\times n}$, where n is the number of classes. To minimize the distance, we use cross-entropy loss $CE(\cdot)$, as follows:

$$L_{CE} = \frac{1}{2}\left[CE(S, Y) + CE(S^{'}, Y)\right], \tag{3}$$

where $S^{'}$ represents the transpose of S.

3 Experiments and Results

In this section, we first describe the datasets, metrics, and implementation details. Then, we present comparison results with SOTA methods, along with an ablation study.

3.1 Datasets and Metrics

The dataset used is collected from four public cohorts: ABIDE [3], ADHD-200 [2], ADNI [7], and OASIS [9]. The dataset comprises 2833 T1w MR images from aged from 6 to 96 years. Specifically, it contains 499 subjects of ASD, 280 subjects of Attention Deficit Hyperactivity Disorder (ADHD), 342 subjects of Early Mild Cognitive Impairment (EMCI), 307 subjects of Late Mild Cognitive Impairment (LMCI), 540 subjects of AD, and 865 Normal Controls (NC). Datasets are split into training and testing sets at an 80:20 ratio. We use Precision, Recall, F1-score, and the area under the curve (AUC) to evaluate the performance.

3.2 Implementation Details

All experiments are conducted on an NVIDIA Tesla V100 GPU with 32 GB of RAM. The Adam optimizer is used with a learning rate of 1×10^{-4} and a batch size of 6. The model is trained for 100 epochs. Brain T1w MR images are first resampled to a uniform voxel spacing of $1.0 \times 1.0 \times 1.0$ mm^3, followed by foreground cropping and resizing to a shape of $(160 \times 192 \times 160)$.

3.3 Performance Comparison

The classification results of Neuro-AMS are shown in Fig. 4. First, we evaluate the performance of Neuro-AMS against three publicly available Vision-Language benchmarks: CLIP [16], BiomedCLIP [24], and XCoOp [1]. While these models originally use 2D Vision Encoders, we ensured a fair comparison by replacing

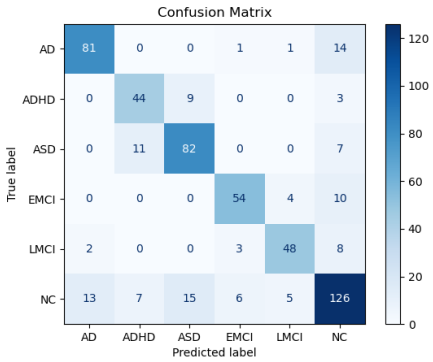

Fig. 4. Confusion matrix of Neuro-AMS for multiple neurological diseases classification.

their 2D Encoder with the same Vision Encoder as used in our model along with the same experimental settings. Table 1 shows the comparisons, and results demonstrate that Neuro-AMS outperforms all these benchmarks.

Second, we compare Neuro-AMS against other SOTA methods designed for specific diseases, as shown in Table 2. To ensure a fair comparison, we selected models which are tested on comparable testing sets. To account for both binary and multi-class classification, we use the recall of each class as the evaluation metric. As shown in Table 2, our method outperforms those specialized models in most classes except for AD (with slightly lower performance). Notably, it achieves a 36.57% improvement for ADHD compared to LEFMs [21]. When compared to 3D-CNN [18], it shows a 10.84% improvement for ASD, with NC remaining similar. Additionally against Hierarchical AD [15], we achieves higher recall scores in EMCI, LMCI, and NC, with improvements of approximately 13%.

Table 1. The quantitative comparison across three vision-language benchmarks, in terms of aAUC, aPRE, aREC, and aF1, where "a" represents the macro average.

	Metric			
Method	aAUC	aPRE	aREC	aF1
CLIP [16]	91.90	74.73	75.44	75.39
BiomedCLIP [24]	91.92	74.53	76.32	75.62
XCoOp [1]	92.85	76.32	77.28	76.83
Neuro-AMS	**93.60**	**78.52**	**79.24**	**79.12**

Table 2. Performance comparison against SOTA models designed for specific diseases, in terms of Recall ("-": denotes no results available).

	Metric					
Method	AD	ADHD	ASD	EMCI	LMCI	NC
$LEFM_s$ [21]	-	42.00	-	-	-	73.26
3D-CNN [18]	-	-	71.16	-	-	**79.34**
Hierarchical AD [15]	**86.70**	-	-	66.30	63.50	62.70
Neuro-AMS	83.50	**78.57**	**82.00**	**79.41**	**78.69**	73.26

3.4 Ablation Results

We conducted a series of ablation experiments to assess the effectiveness of various modules within Neuro-AMS. The quantitative results of these experiments are summarized in Table 3.

As observed, ablation studies show that both the semantic-assisted method and age-aware strategy improve performance over the baseline by 18.2% and 10.7%, respectively. A performance increase of 1.3% observed when integrating the age-aware strategy into the semantic-assisted framework further validates its contribution. Furthermore, in the knowledge-integrated ablation experiment, incorporating medical knowledge resulted in a 2.5% average performance improvement, highlighting the value of integrating medical knowledge to enhance feature discrimination compared to using disease labels alone.

Table 3. Ablation experiments on the multi-disease dataset comprising four public cohorts, in terms of aAUC, aPRE, aREC, and aF1, where "a" represents the macro average vector<vector<int» generateMatrix(int n)

Baseline	Text-assistance		Age-aware	aAUC	aPRE	aREC	aF1
	Label	Knowledge					
✓	✗	✗	✗	64.61	52.35	68.52	58.43
✓	✗	✗	✓	81.35	70.51	71.64	63.75
✓	✓	✗	✗	91.90	74.73	75.44	75.39
✓	✓	✗	✓	93.10	75.81	76.92	75.60
✓	✓	✓	✓	**93.60**	**78.52**	**79.24**	**79.12**

4 Conclusion

We have introduced Neuro-AMS, a semantic-assisted framework for diagnosing multiple neurological diseases. The core advantage of our method lies in its ability to incorporate the inherent semantics of medical conditions and the influence

of age, significantly enhancing the model's capacity to differentiate among neurological diseases. Experimental results demonstrate that our approach not only outperforms other SOTA methods but also demonstrates a marked improvement in diagnostic accuracy and robustness. Neuro-AMS shows great promise for advancing clinical applications by offering more accurate and interpretable diagnosis of multiple neurological diseases, paving the way for more robust and interpretable diagnoses. Moving forward, we plan to expand the framework to include a broader range of neurological diseases and incorporate more diverse data modalities to further enhance its diagnostic capabilities.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 82441023, U23A20295, 62131015, 82394432), the China Ministry of Science and Technology (S20240085, STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040 001), and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bie, Y., Luo, L., Chen, Z., Chen, H.: Xcoop: Explainable prompt learning for computer-aided diagnosis via concept-guided context optimization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 773–783. Springer (2024)
2. consortium, A..: The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in systems neuroscience **6**, 62 (2012)
3. Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry **19**(6), 659–667 (2014)
4. Dufumier, B., Gori, P., Victor, J., Grigis, A., Wessa, M., Brambilla, P., Favre, P., Polosan, M., Mcdonald, C., Piguet, C.M., et al.: Contrastive learning with continuous proxy meta-data for 3d mri classification. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. pp. 58–68. Springer (2021)
5. Fan, J., Cao, X., Wang, Q., Yap, P.T., Shen, D.: Adversarial learning for mono-or multi-modal registration. Medical image analysis **58**, 101545 (2019)
6. Iqbal, M.S., Heyat, M.B.B., Parveen, S., Hayat, M.A.B., Roshanzamir, M., Alizadehsani, R., Akhtar, F., Sayeed, E., Hussain, S., Hussein, H.S., et al.: Progress and trends in neurological disorders research based on deep learning. Computerized Medical Imaging and Graphics **116**, 102400 (2024)

7. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
8. Kowalska, M., Owecki, M., Prendecki, M., Wize, K., Nowakowska, J., Kozubski, W., Lianeri, M., Dorszewska, J.: Aging and neurological diseases. In: Senescence-physiology or pathology. IntechOpen (2017)
9. LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., et al.: Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. medrxiv pp. 2019–12 (2019)
10. Liu, M., Zhang, D., Adeli, E., Shen, D.: Inherent structure-based multiview learning with multitemplate feature representation for alzheimer's disease diagnosis. IEEE Transactions on Biomedical Engineering **63**(7), 1473–1482 (2015)
11. Liu, Y., Liu, M., Zhang, Y., Guan, Y., Guo, Q., Xie, F., Shen, D.: Amyloid-β deposition prediction with large language model driven and task oriented learning of brain functional networks. IEEE Transactions on Medical Imaging (2025)
12. Liu, Y., Liu, M., Zhang, Y., Sun, K., Shen, D.: A progressive single-modality to multi-modality classification framework for alzheimers disease sub-type diagnosis. In: International Workshop on Machine Learning in Clinical Neuroimaging. pp. 123–133. Springer (2025)
13. Lohani, D.C., Rana, B.: Adhd diagnosis using structural brain mri and personal characteristic data with machine learning framework. Psychiatry Research: Neuroimaging **334**, 111689 (2023)
14. Okoye, C., Obialo-Ibeawuchi, C.M., Obajeun, O.A., Sarwar, S., Tawfik, C., Waleed, M.S., Wasim, A.U., Mohamoud, I., Afolayan, A.Y., Mbaezue, R.N.: Early diagnosis of autism spectrum disorder: a review and analysis of the risks and benefits. Cureus **15**(8) (2023)
15. Qin, Y., Cui, J., Ge, X., Tian, Y., Han, H., Fan, Z., Liu, L., Luo, Y., Yu, H.: Hierarchical multi-class alzheimers disease diagnostic framework using imaging and clinical features. Frontiers in Aging Neuroscience **14**, 935055 (2022)
16. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
17. Rao, Y.L., Ganaraja, B., Murlimanju, B., Joy, T., Krishnamurthy, A., Agrawal, A.: Hippocampus and its involvement in alzheimers disease: a review. 3 Biotech **12**(2), 55 (2022)
18. Shahamat, H., Abadeh, M.S.: Brain mri analysis using a deep learning based evolutionary approach. Neural Networks **126**, 218–234 (2020)
19. Steinmetz, J.D., Seeher, K.M., Schiess, N., Nichols, E., Cao, B., Servili, C., Cavallera, V., Cousin, E., Hagins, H., Moberg, M.E., et al.: Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the global burden of disease study 2021. The Lancet Neurology **23**(4), 344–381 (2024)
20. Teng, L., Zhao, Z., Huang, J., Cao, Z., Meng, R., Shi, F., Shen, D.: Knowledge-guided prompt learning for lifespan brain mr image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 238–248. Springer (2024)

21. Tian, L., Zheng, H., Zhang, K., Qiu, J., Song, X., Li, S., Zeng, Z., Ran, B., Deng, X., Cai, J.: Structural or/and functional mri-based machine learning techniques for attention-deficit/hyperactivity disorder diagnosis: A systematic review and meta-analysis. Journal of affective disorders (2024)
22. Windsor, R., Jamaludin, A., Kadir, T., Zisserman, A.: Vision-language modelling for radiological imaging and reports in the low data regime. arXiv preprint arXiv:2303.17644 (2023)
23. Yao, H., Zhang, R., Xu, C.: Tcp: Textual-based class-aware prompt tuning for visual-language model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23438–23448 (2024)
24. Zhang, S., Xu, Y., Usuyama, N., Xu, H., Bagga, J., Tinn, R., Preston, S., Rao, R., Wei, M., Valluri, N., et al.: Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv preprint arXiv:2303.00915 (2023)
25. Zhang, X., Wu, C., Zhang, Y., Xie, W., Wang, Y.: Knowledge-enhanced visual-language pre-training on chest radiology images. Nature Communications **14**(1), 4542 (2023)
26. Zhao, Z., Liu, Y., Wu, H., Wang, M., Li, Y., Wang, S., Teng, L., Liu, D., Cui, Z., Wang, Q., et al.: Clip in medical imaging: A comprehensive survey. arXiv preprint arXiv:2312.07353 (2023)