Progression-Aware Generative Model Enhancing Baseline Visit Prediction of Early Alzheimer's Disease

Xingyu Gao[1], Runyi Guo[2], Zhe Zhao[1]

[1] College of Biological Science and Medical Engineering, Donghua University, Shanghai, China
zhezhao@dhu.edu.cn
[2] College of Physics, Donghua University, Shanghai, China

Abstract. Benefiting from longitudinal pair-wise brain ^{18}F-fluorodeoxyglucose (^{18}F-FDG) positron emission tomography (PET) images, disease progression characterized by the generative model may assist the baseline visit prediction of early Alzheimer's Disease. However, most existing methods focused on diagnosing disease from single-timepoint scans or a simple stacking of sequential images, which ignore the importance of disease progression and are not in line with actual clinical scenarios. Moreover, decoupling the low-level disease representations is quite challenging for similar changes between normal aging and neurodegenerative changing. In this paper, we propose a classifier induced generative model to generate the next-timepoint brain images. Then, we design a statistical prior knowledge vision transformer to extract features from the generated next-timepoint images for disease diagnosis. The main contribution is to build a disease progression model that can effectively improve diagnosis performance from single-timepoint images. Meanwhile, we provide pixel-level disease representations for explanation. Experiments on ADNI datasets demonstrate that our method outperforms other state-of-the-art techniques.

Keywords: Alzheimer's Disease, PET, Image Generation, Disease Diagnosis.

1 Introduction

Alzheimer's Disease (AD) is a progressive and irreversible neurodegenerative brain disorder, which is mainly characterized by memory impairment and cognitive decline [1]. With the increase of global aging population, the number of individuals affected by dementia is projected to 100 million by 2050 [2]. AD as the most common form of dementia, accounts for approximately 60-70% of all cases. Thus, accurate diagnosis of early AD is critical and pressing. Positron emission tomography (PET) images can capture functional changes caused by neurodegenerative disorder before structures atrophy, which is widely used in clinical diagnosis. In recent years, there are many deep learning methods that concentrate on AD diagnosis with PET scans [3-6].

Constructing a diagnostic model on single-timepoint cross-sectional images is the most straightforward and widely applicable approach. Ding et al. [4] took an Inception V3 on ADNI dataset for distinguishing AD versus all other cases and achieved the AUC

of 92%, which demonstrates that the deep learning method can perform accurate and robust diagnosis of AD from FDG-PET images. Tau-PET images can reflect tau accumulation and pathological changes of AD, a convolutional neural network (CNN) integrating tau-PET scans and demographic information is proposed, offering an effective approach to enhance the classification between mild cognitive impairment (MCI) and cognitive normal (CN) [5]. Taking advantage of the transfer learning technique [6], a 2D slice-level CNN is pre-trained on the ImageNet dataset and fine-tuned with the ADNI dataset. This model struggles to fast converge with 50% dropout, achieving a remarkable testing accuracy of 91.43% for the classification of healthy versus AD patients. The methods mentioned above are prone to exploit shortcut patterns on medical images. Although they have shown promising performance, it is still challenging to meet actual clinical requirements for the lack of expert prior knowledge and explicit explanations.

Embedding expert knowledge into image-wise classification models requires extensive medical descriptions or handcrafted annotations, which comes at a substantial cost. Finding biomarkers and leveraging statistical information are both helpful and explainable for disease prediction [7-9]. Jiang et al. [7] proposed an anatomy-aware gating network for Alzheimer's disease prediction, which explicitly extracts features from anatomical regions using an anatomy-aware squeeze-and-excite operation. It highlights the brain regions of importance, improving transparency and assisting clinicians in identifying areas for further examination. In terms of statistical analysis, t-test, z-score, Wilcoxon rank sum test, Chi-square test and binomial test are used to assess the effect of individual PET biomarkers. Then, a univariate logistic regression model is adopted to accurately predict the presence of cognitive decline and neuropathological outcome [8]. To detect the individualized pathological changes, a brain status transferring generative adversarial network is proposed [9]. This network generates corresponding healthy brain images from patient data. Then, by computing residuals between the real and generated images, the severity of the disease can be effectively quantified. These methods leverage biomarkers and statistical prior knowledge to enhance the performance and interpretability of classification models, thereby advancing the development of assisted diagnosis.

Compared with cross-sectional scans, longitudinal medical sequences can characterize the disease progression. However, most methods focus on statistical analysis alone, that are not integrated with diagnostic models [10,11]. Recently, many deep learning-based methods have been proposed to learn the disease representations from temporal images [12,13]. These methods stack temporal images together for disease prediction, which does not align with real-world clinical scenarios. In practice, temporal medical data is very limited because patients fail to follow up or seek treatment at different medical centers. Making good use of existing temporal data to assist single-timepoint diagnoses may achieve promising performance.

In this study, we propose a classifier-induced generative model to generate the next-timepoint brain images from baseline visit scans. Then, the disease progression patterns from temporal images are used to enhance the predictive performance on baseline visit scans. Besides, we design a prior knowledge vision transformer to extract features from the generated next-timepoint images for disease diagnosis. There are two reasons for

using PET scans as study data instead of magnetic resonance imaging (MRI). First, PET imaging with amyloid-specific tracers is the gold standard for assessing AD, and many studies have demonstrated that diagnostic performance is superior when using PET images [14,15]. Second, PET images exhibit better data consistency compared with MRI data. Therefore, no additional image preprocessing operations are needed for paired temporal PET images.

In summary, our main contributions are as follows:

- We construct a classifier-induced generative model from limited temporal brain images to generate the disease progression for the baseline visit single-timepoint images.
- We design a prior-knowledge vision transformer to extract features from the generated next-timepoint images for disease diagnosis. Pixel-level disease representations is performed for explanation.

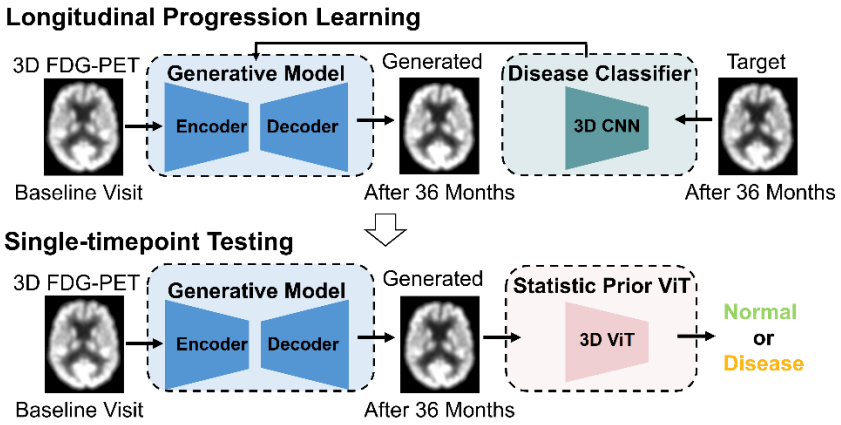

Fig. 1. Overview of our proposed method. Longitudinal progression learning module learns the normal or disease progression from paired FDG-PET. Single-timepoint testing extracts the representation from the generated next-timepoint images.

2 Methodology

2.1 Overview

Developing diagnostic models for cross-sectional images have reached a performance plateau due to the inherent properties of source data. To enhance the diagnostic performance on baseline visit images, we aim to introduce additional progression information from longitudinal images. Specifically, the baseline visit group include cognitive normals (CN) and progressive mild cognitive impairments (pMCI). At the next timepoint, CN subjects maintain their CN status, while pMCI patients progress to Alzheimer's disease (AD). It is obvious that the classification accuracy for AD vs. CN is higher than that for pMCI vs. CN when using the same method. Inspired by this, we propose a classifier-induced generative model and a prior knowledge-based vision transformer

network. An overview of our proposed method is shown in Fig.1. Detailed description of each module is provided in the following subsections.

2.2 Classifier-induced generative model for progression learning

Generative models excel in image-to-image translation especially in generating the diversity of natural scenes. Generating brain images of individuals with neurodegenerative diseases or cognitive normals presents a significant challenge due to the similar changes. For the task of cross-timepoint generation, we construct a classifier-induced generative model which consists of two modules: 1) generative adversarial network (GAN) and 2) disease classification network. To demonstrate the effectiveness of our thinking, we utilize a basic GAN with a U-Net generator and a lightweight 3D CNN discriminator.

Let X_{BL} and X_{NT} be the baseline visit and the next-timepoint image, respectively. The loss function of the generator and discriminator are formulated as:

$$L_G = E\left[log\left(1-D\left(G\left(X_{BL}\right)\right)\right)\right] \tag{1}$$

$$L_D = E\left[log\left(1-D\left(X_{NT}\right)\right)\right] + E\left[log\left(D\left(G\left(X_{BL}\right)\right)\right)\right] \tag{2}$$

where G denotes the generator and D denotes the discriminator. This adversarial training process ensures that the generated images adhere to the PET modality.

To improve the ability of the generative model, we introduce pixel-level supervision between the generated $\hat{X}_{NT}$ and the real X_{NT} such as the index of L_1 and structural similarity index matrix ($SSIM$) which can be defined as:

$$L_1 = \left\| X_{NT} - X_{NT} \right\|_1 \tag{3}$$

$$L_{SSIM} = \frac{1}{|M|}\sum SSIM\left(X_{NT}, X_{NT}\right) \tag{4}$$

$$SSIM\left(X_{NT}, X_{NT}\right) = \frac{\left(2\mu_{X_{NT}}\mu_{X_{NT}} + C_1\right)\left(2\sigma_{X_{NT}X_{NT}} + C_2\right)}{\left(\mu^2_{X_{NT}} + \mu^2_{X_{NT}} + C_1\right)\left(\sigma^2_{X_{NT}} + \sigma^2_{X_{NT}} + C_2\right)} \tag{5}$$

where M, μ , σ are the number of local windows, mean value and the standard deviation of the image. C_1 and C_2 are constants which are introduced to prevent the zero denominator.

Above basic GAN can generate a global description of the target image but not the disease progression for individuals. An expert model (i.e. DenseNet-18 [16]) learned knowledge from target images can effectively assist the generative model. In our study, we train a 3D CNN from target images as the expert model. The loss function of expert model is computed as:

$$L_C = E\left[-\mathrm{y}\,log\left(D_c\left(G\left(X_{BL}\right)\right)\right) - (1-\mathrm{y})\,log\left(1-D_c\left(G\left(X_{BL}\right)\right)\right)\right] \tag{6}$$

where D_c denotes the classifier. And the performance of this module will show in the section of experiments and results.

2.3 Prior knowledge-based vision transformer for disease prediction

Using the generated next-timepoint images, we propose a prior-knowledge vision transformer network as shown in Fig. 2. First, we compute the residuals between the real baseline visit image and next-timepoint image cross all subjects. Then, a pixel-level merging operation, followed by a normalization, is performed. As a result, we obtain a brain atlas that is highly correlated with the disease. The brain atlas is used as prior knowledge for our classification network. For image representation, we leverage DenseNet-18 [16] as the backbone due to its acknowledged capability of feature extraction. In terms of the high-level image representation, we introduce the transformer to facilitate information interaction. Finally, a fully connected layer with Softmax activation function is used to predict the subject label.

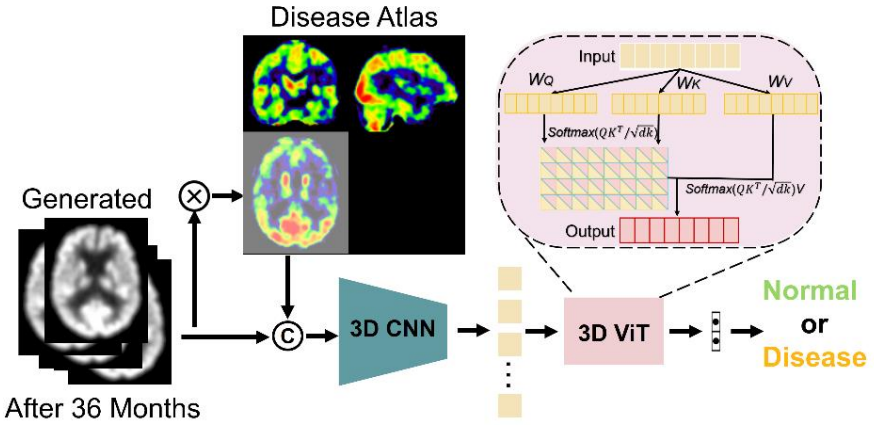

Fig. 2. The architecture of our prior-knowledge vision transformer network (Prior ViT).

Mathematically, let the generated next-timepoint image as $\hat{X}_{NT}$, the matrix of brain atlas is denoted as B. Thus, the high-level image representation R can be formulated as $R = f(\hat{X}_{NT}, \hat{X}_{NT} * B)$, where f is the feature mapping function. Then, the transformer is used to model the interactions of high-level features. Then, a convolution operation is applied to R, projecting it to Q, K, and V, respectively. Through the transformer, the output can be computed as:

$$J = \text{Softmax}\left(QK^T / \sqrt{d}\right)V \tag{7}$$

where d is the dimension of the Q, K, V. After the fully connected layer and Softmax activation, we get the predicted label $\hat{y}$. Finally, the loss function is denoted as:

$$L_{cls} = -\sum_{i=1}^{N}[y * log\left(y\right) + \left(1-y\right) * log\left(1-y\right)] \tag{8}$$

3 Materials

In this work, the longitudinal PET images are from ADNI dataset. Timepoints contain baseline visit and follow up after 36 months. 5-fold cross-validation are used to test the performance of our method. In the dataset, there are total 280 subjects (including 88 CN subjects and 192 patients with pMCI who converted to Alzheimer's disease within 36 months). To maintain a balanced distribution across classes in the test set, we randomly select an equal number of samples from the pMCI group and CN group. The image pre-processing procedures include skull-stripping and linear registration using the FSL software. Then, the post-process images are normalized to values between 0 and 1. Finally, the images are downsampled to 76 × 94 × 76 as the inputs for efficiency.

4 Experiments and Results

4.1 Implementation

Our proposed methods were developed on Pytorch framework within Ubuntu 24.04 LTS. The generative model was trained with batch size = 2, learning rate = 1e-4, and epoch = 300. Adam optimizer was used to update the learning rate. To mitigate the overfitting risk, L_2 weight decay was set to 1e-4. At the first training stage (i.e., iterations < 100), the loss function = $L_G + L_1 + L_{SSIM}$. And then the loss function was set to $L_G + L_1 + L_{SSIM} + L_D$ when iterations greater than 100 and less than 200. In the last 100 epochs, the loss function = $L_G + L_1 + L_{SSIM} + L_D + L_C$. For the predictive model, the batch size, learning rate, and epoch were set to 6, 1e-3, and 60, respectively. The optimizer settings were the same as the generative model. Above two models were accelerated by 1 Nvidia GeForce RTX 4090 D.

4.2 Results of Image Generation

To evaluate the effectiveness of our classifier-induced generative model, we show 4 typical results on the test set in Fig. 3. First, we calculate the residual between the real 36-month (M36) image and the baseline visit (BL) image to characterize disease progression. Then, we compute the residual between the generated 36-month (M36_Gen) image and the baseline visit (BL) image. From the first two results (i.e., 128_S_0227 and 072_S_1211), we can see that the residual images from each patient are similar, which indicates that our classifier-induced generative model can map the disease progression. From the last two results (i.e., 037_S_0327 and 035_S_0048), there are several highlight regions on the residual images, that are normal changes caused by aging. Moreover, the positions of the highlighted regions differ among test samples, proving that our generative model can achieve an individualized diagnosis.

Second, we use the expert model (i.e., the classification model mentioned in section 2.2) to quantify the quality of the generated PET images. For comparative analysis, we train the expert model on cross-section data of BL, real M36, and generated M36, respectively. The training and testing sets include the same subjects. Table 1 shows the classification results.

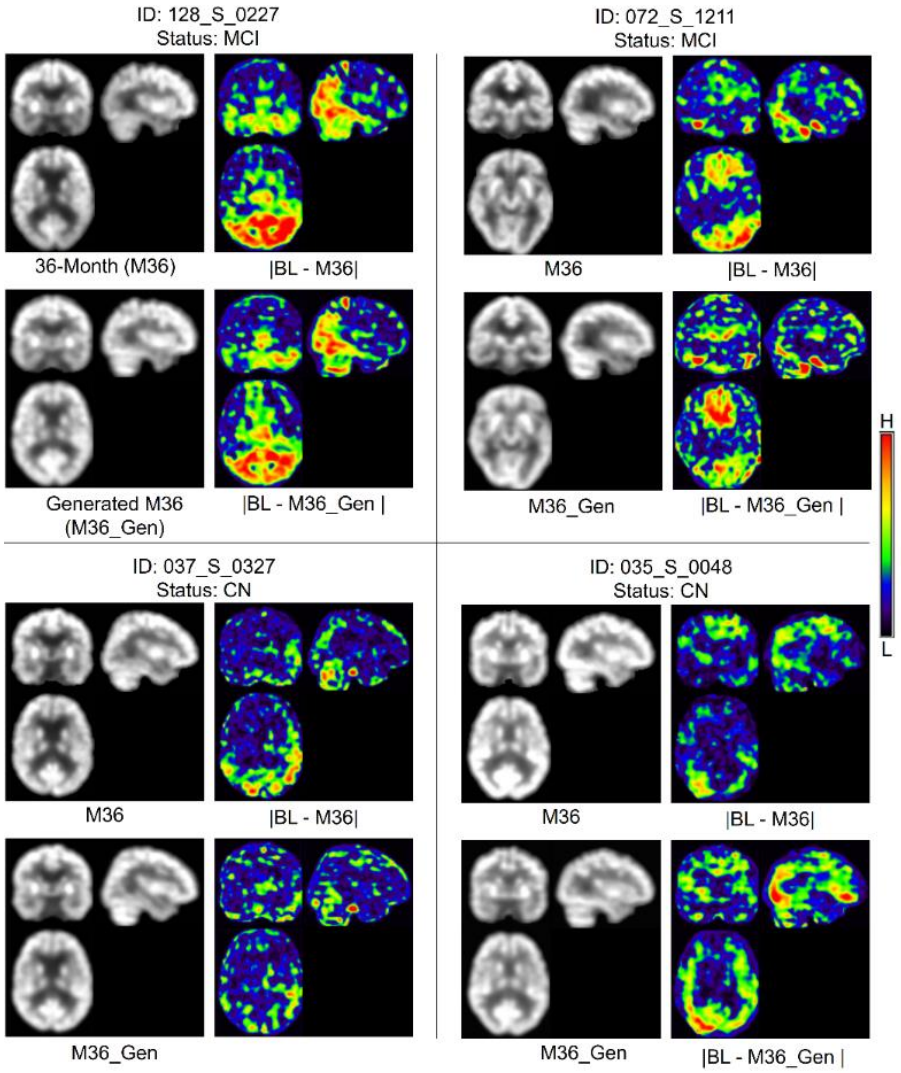

Fig. 3. Illustration of real 36-month (M36) PET images, their corresponding generated results, and residuals between baseline visit (BL) image and M36/M36_Gen image.

Table 1. Performance of expert model [16] on different timepoint data.

Timepoint Data	pMCI *vs* CN (%)				
	AUC	Accuracy	Sensitivity	Specificity	F1-score
BL	74.2	73.3	73.3	73.3	73.3
Real M36	**82.7**	**83.3**	**80.0**	**86.7**	**82.8**
Generated M36	80.3	80.0	**80.0**	80.0	80.0

From Table 1, we make two observations. First, the classification performance on M36 timepoint images is better than that on BL images. The above results indicate that the disease characteristics will become more distinct over time. Second, the classification results on Generated M36 images are greater than that on BL images, which demonstrates the effectiveness of our classifier-induced generative model for cross-timepoint image generation. In this section, we do not introduce architectural innovations to the model, as the primary objective is to validate the effectiveness of the proposed paradigm.

4.3 Results of Disease Prediction

In this section, we compare our Prior ViT with other state-of-the-art classification methods, including (1) a multi-stream convolutional neural network [17]; (2) the expert model [16]; and (3) the Prior ViT without disease atlas (denoted as Our prior ViT w/o DA). These methods are trained and tested using the baseline visit images. Specifically, the first approach is trained and tested on original images, while the last three are trained and tested on synthetically generated 36-month images from original images. Table 2 compares the classification performance of four different methods.

Table 2. Classification results achieved by four different methods.

Method	pMCI *vs* CN (%)				
	AUC	Accuracy	Sensitivity	Specificity	F1-score
Ashtari-Majla et.al., 2022	72.4	66.7	66.7	66.7	66.7
Huang et.al., 2017	80.3	80.0	**80.0**	80.0	80.0
Our prior ViT (w/o DA)	81.7	80.0	**80.0**	80.0	80.0
Our prior ViT	**83.5**	**83.3**	**80.0**	**86.6**	**82.7**

From Table 2, we can observe that the methods trained and tested on the generated 36-month images perform better than those trained and tested on original images. Additionally, the performance of our prior ViT without the assistance of disease atlas has decreased. Our prior ViT achieves the best performance in classification of pMCI vs CN, indicating the effectiveness of our proposed method.

5 Conclusion

In this paper, we propose a classifier-induced generative model and a prior-knowledge vision transformer model. The classifier-induced generative model learns to map the 36-month image from the baseline visit image. As a result, the generated 36-month images can not only maintain global similarity with the target image but also capture the disease progression. Then, disease prediction is performed by the prior-knowledge vision transformer model. Experiments on ADNI dataset demonstrate the effectiveness of our proposed method.

Acknowledgments. This research was supported in part by the Fundamental Research Funds for the Central Universities (2232025D-38, 2232024D-33).

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

References

1. Wang J., et al.: PET molecular imaging for pathophysiological visualization in Alzheimer's disease. Eur J Nucl Med Mol Imaging. **50**(3), 765–783 (2023).
2. Nichols E., et al.: Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. The Lancet Public Health. **7**(2), e105-e125 (2022).
3. Xue C., et al.: AI-based differential diagnosis of dementia etiologies on multimodal data. Nat Med. **30**(10), 2977–2989 (2024).
4. Ding Y., et al.: A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using ^{18}F-FDG PET of the Brain. Radiology. **290**(2), 456–464, (2019).
5. Park S. W., et al.: Deep learning application for the classification of Alzheimer's disease using 18F-flortaucipir (AV-1451) tau positron emission tomography. Sci Rep. **13**(1), 8096 (2023).
6. Bastos J., et al.: Deep Learning for Diagnosis of Alzheimer's Disease with FDG-PET Neuroimaging. in Pattern Recognition and Image Analysis, pp. 95–107. Springer, Cham (2022).
7. Jiang H., et al.: Anatomy-Aware Gating Network for Explainable Alzheimer's Disease Diagnosis. in MICCAI 2024, vol. 15005, pp. 90–100. Springer, Cham (2024).
8. Long J. M., et al.: Preclinical Alzheimer's disease biomarkers accurately predict cognitive and neuropathological outcomes. Brain. **145**(12), 4506–4518 (2022).
9. Gao X., et al.: Brain Status Transferring Generative Adversarial Network for Decoding Individualized Atrophy in Alzheimer's Disease. IEEE J. Biomed. Health Inform. **27**(10), 4961–4970 (2023).
10. Aberathne I., et al.: Detection of Alzheimer's disease onset using MRI and PET neuroimaging: longitudinal data analysis and machine learning. Neural Regen Res. **18**(10), 2134 (2023).
11. Wearn A.: Longitudinal changes in hippocampal texture from healthy aging to Alzheimer's disease. Brain Communications. **5**(4), fcad195 (2023)
12. Ebrahimi A., et al.: Deep sequence modelling for Alzheimer's disease detection using MRI. Computers in Biology and Medicine. **134**, 104537, (2021).
13. Fang C., et al.: Deep learning for predicting COVID-19 malignant progression. Medical Image Analysis. **72**, 102096 (2021).
14. Fan S., et al.: Classification of amyloid positivity in PET imaging using end-to-end deep learning: a multi-cohort, multi-tracer analysis. Alzheimer's & Dementia. **19**(S2), e067388 (2023).
15. Gao X., et al.: Task-Induced Pyramid and Attention GAN for Multimodal Brain Image Imputation and Classification in Alzheimer's Disease. IEEE J. Biomed. Health Inform. **26**(1), 36–43 (2022).
16. Huang G., et al.: Densely Connected Convolutional Networks. in CVPR. pp. 2261-2269 (2017).

17. Ashtari-Majla M., et al.: A multi-stream convolutional neural network for classification of progressive MCI in Alzheimer's disease using structural MRI images. IEEE J. Biomed. Health Inform. **26**(8), 3918–3926 (2022).