Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 14, No. 4, 2026, 621-632.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Biological Brain Age Estimation Using Multi-Task Self-Supervision and Pretrained Deep Neural Networks

Zahrasadat Sajjadi, Soheil Hamzebeigi and Mohsen Soryani*

School of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Article Info

Article History:
Received 20 April 2026
Revised 05 July 2026
Accepted 04 August 2026

DOI:10.22044/jadm.2026.17592.2904

Keywords:
Deep Learning, DINO, Multi-Task Learning, Pretraining, Transfer Learning

Corresponding author: soryani@iust.ac.ir (M. Soryani).

Abstract

As the global population ages, reliable methods for assessing brain health and age-related changes are increasingly important. Brain age is a promising biomarker of brain health, and machine-learning methods have enabled its estimation from neuroimaging data. However, effective training strategies are required for accurate brain age estimation. This study proposes a two-dimensional convolutional neural network (2D CNN)-based multi-task framework for estimating brain age from magnetic resonance imaging (MRI) scans in the ADNI dataset. The framework uses VGG-16 and U-Net encoder backbones with separate heads for age regression and image-rotation classification. We evaluated the effects of several pretraining strategies, including self-supervised pretraining using the DINO framework and supervised pretraining through brain-tumor segmentation. The best-performing configuration, consisting of a DINO-pretrained VGG-16 backbone and a multi-task prediction head, achieved a mean absolute error (MAE) of 3.27 years, which is competitive with previously reported methods. The results indicate that combining transfer learning with multi-task learning improved performance relative to the corresponding single-task models, suggesting that this combination supports the learning of richer and more generalizable feature representations.

1. Introduction and Related Works

The structure of the human brain changes with age, and these changes can be observed using magnetic resonance imaging (MRI). Neuroimaging data, such as MRI scans, can be analyzed to estimate an individual's biological brain age. This estimate can provide insights into overall brain health and serve as a biomarker for disease [1], [2].

Machine learning techniques can accurately estimate the biological age of the brain, thereby quantifying the difference between an individual's chronological age and the estimated biological age. This difference, expressed through the Absolute Error (AE), determines whether an individual's brain appears younger or older than their chronological age. This information is beneficial for understanding healthy brain aging and may support early detection and intervention in conditions such as Alzheimer's disease, mild cognitive impairment, and schizophrenia [3].

While traditional machine learning approaches—including support vector regression, relevance vector regression, and Gaussian process regression—have been widely used to estimate brain age from MRI[1], recent studies increasingly favor deep learning methods for predicting biological age, reflecting a shift toward more complex, data-driven models.

In 2017, Cole et al.[4] were among the first to apply deep learning to estimate brain age. Their objective was to evaluate estimated brain age as a biomarker and to demonstrate the superior effectiveness of

deep learning models over traditional machine learning models in brain age estimation. Using a convolutional model and brain gray matter data obtained from T1-weighted MRI scans, they achieved an MAE of 4.16 years. Furthermore, they showed that estimated brain age is a reliable and genetically influenced phenotype that can serve as a biomarker of brain aging.

In 2019, Jonsson et al.[5] developed a new approach to estimate brain age using T1-weighted MRI data. Their 3D convolutional model, based on ResNet-50[6], made two significant improvements. First, the model incorporated sex and scanner type as input variables. Second, brain age estimation was performed using deformation features derived from Jacobian maps. The authors demonstrated that fine-tuning a pretrained model significantly enhanced prediction accuracy.

Bashyam et al.[7] (2021) explored brain aging and pathology within a unified modeling framework. Their study used T1-weighted MRI data from 14,468 healthy participants, ranging in age from 3 to 95 years, acquired across diverse scanners and demographic backgrounds. The authors developed a convolutional DeepBrainNet model to estimate brain age with minimal preprocessing, achieving an MAE of 3.702 years. Furthermore, they found that models that perform well in estimating biological age may not be optimal for diagnosing diseases such as Alzheimer's and schizophrenia.

Deep learning models for brain age estimation often face limitations due to the scarcity of large-scale training datasets and high computational and memory requirements. To overcome these challenges, Peng et al.[8] introduced the Simple Fully Convolutional Neural Network (SFCN) in 2021, a streamlined architecture designed for efficiency and accuracy. This model used fewer parameters than other popular 3D deep networks. The SFCN successfully estimated brain age using structural T1 MRI data, achieving an MAE of 2.14 years for brain age estimation and 99.5% accuracy in sex classification on UK Biobank[9] data, which comprised 14,503 samples aged between 44 and 80 years. The results suggest that lightweight models like SFCN may perform as well as, or even better than, larger models in medical imaging contexts, calling into question the widely held belief that larger architectures necessarily yield superior results, as is often the case in natural image analysis.

Hwang et al.[1] conducted a study in 2021 to estimate brain age using T2-weighted axial images acquired during routine clinical exams. To estimate human brain age, the authors developed a predictive model based on a deep convolutional neural network architecture. The model was evaluated using 10-fold cross-validation. Performance was assessed using MAE; the five best folds each achieved an MAE below 5 years. The model reported an MAE of 4.22 years on an internal test set of 270 samples and 9.96 years on an external test set of 560 samples. While the internal data were gathered over five years from a single hospital in Seoul, the external data originated from a publicly available source. Disparities in imaging parameters and ethnicity likely contributed to the observed performance differences, highlighting the importance of dataset standardization and broader validation in future studies.

In 2021, Hedderich et al.[10] used a classical machine learning method, relevance vector regression (RVR), with a linear kernel. Using MRI data from 648 healthy individuals aged 11 to 70 years, the researchers trained their model to explore the potential impact of premature birth on brain age. Furthermore, the study investigated how brain age relates to perinatal variables and long-term developmental outcomes both physical and cognitive using a longitudinal cohort of 101 young adults born prematurely and 111 full-term controls, followed from birth through their third decade. They found increased brain age in premature-born adults compared to full-term controls, which was associated with low gestational age and low birth weight but not with total intracranial volume or full-scale IQ. The results demonstrated elevated brain age in premature-born adults, suggesting potentially accelerated brain aging in premature-born individuals. In 2023, Pacheco et al.[2] investigated whether pretraining on a brain-related task, specifically tumor segmentation, could improve brain age estimation using two backbone architectures: U-Net[11] and ResUNet[12]. For brain age estimation, the researchers employed T1-weighted MRI scans of healthy individuals from the ADNI dataset for training, validation, and testing. Additionally, they used the BraTS 2020 dataset for brain-tumor segmentation. The findings indicated that using the U-Net backbone with the proposed pretraining approach yielded the best results, with an MAE of 3.07 years.

The main contribution of this study lies in the evaluation of pretraining and multi-task learning strategies rather than the development of a new CNN architecture. We examined two pretraining pathways for 2D brain age estimation: self-supervised DINO pretraining on brain MRI and supervised brain-domain pretraining through tumor segmentation. Within each pathway, we compared single-task age regression with a multi-

task framework that jointly learns age regression and rotation classification. We also compared rotation prediction with conventional rotation-based data augmentation to determine whether the auxiliary objective provides benefits beyond increased image variability. Through these comparisons, the study investigates how different forms of prior knowledge and auxiliary supervision can be used to improve brain age estimation when labeled neuroimaging data are limited.

This study estimates biological brain age using a multi-task self-supervised model. In addition to the downstream brain age estimation task, the proposed multi-task model incorporates image-rotation classification as an auxiliary task to improve feature learning. We also evaluate the effects of self-supervised and supervised pretraining on the proposed multi-task model. The following six experimental scenarios were considered in this study:

a) Single-task brain age estimation without additional pretraining using a VGG-16 backbone.
b) Multi-task brain age estimation and image rotation classification without additional pretraining using a VGG-16 backbone.
c) Multi-task brain age estimation and image rotation classification without pretraining (randomly initialized U-Net encoder backbone).
d) Multi-task brain age estimation and image rotation classification using a U-Net encoder pretrained for brain tumor segmentation.
e) Single-task brain age estimation using a VGG-16 backbone pretrained with the self-supervised DINO framework on the BraTS 2020 dataset.
f) Multi-task brain age estimation and image rotation classification using a VGG-16 backbone pretrained with the self-supervised DINO framework on the BraTS 2020 dataset.

2. Materials and Methods

2.1. Model Architectures

To implement the six different scenarios (a)-(f) mentioned in the previous section, we used two 2D convolutional networks: VGG-16[13] and U-Net.

To ensure a fair evaluation, the six scenarios were not interpreted solely through an overall comparison. Instead, they were compared in controlled pairs, with the main architectural and training conditions kept constant within each pair. Specifically, scenarios (a) and (b) were compared to evaluate the effect of multi-task learning under the same VGG-16 backbone and initialization. Scenarios (e) and (f) were compared to assess the effect of multi-task learning when both models used a DINO-pretrained VGG-16 backbone. Scenarios (c) and (d) were compared to determine the effect of tumor-segmentation pretraining while using the same U-Net encoder and prediction head. In addition, scenarios (b) and (f) were compared to evaluate the effect of DINO pretraining in the multi-task VGG-16 setting, whereas scenarios (a) and (e) were compared to assess its effect in the single-task VGG-16 setting. Therefore, the main conclusions were drawn from these pairwise comparisons rather than from direct comparisons across models with different architectures and parameter counts. We selected VGG-16 and U-Net because they were well suited to the two pretraining strategies examined in this study, rather than because they offered the lowest computational cost. VGG-16 has a simple and consistent structure based on stacked 3 × 3 convolutional layers, which made it straightforward to transfer the DINO-pretrained weights and compare single-task and multi-task heads using the same backbone. U-Net was used for tumor-segmentation pretraining because its encoder can learn from pixel-level supervision and can later be reused for age estimation after removing the decoder. We acknowledge that newer lightweight models, such as SFCN, may provide a better balance between accuracy and computational efficiency. These models were not included because our main aim was to study the effects of pretraining and an auxiliary task, rather than to compare a wide range of architectures. Following the VGG-16 backbone, we used a multi-task prediction head comprising two fully connected layers with 4096 and 2048 units, respectively. These shared layers were followed by two task-specific output layers: a single-node regression layer for brain age estimation and a four-node classification layer for image-rotation prediction. The complete architecture is illustrated in Figure 1.

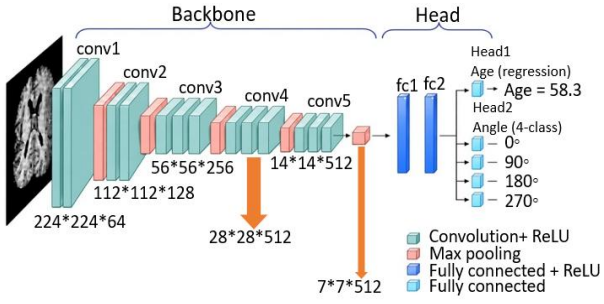

Figure 1. Structure of the proposed model with the VGG-16 backbone and multi-task head.

The second architecture was based on U-Net. U-Net is widely used for image segmentation and consists of an encoder, a bottleneck, and a decoder. The complete architecture is shown in Figure 2.
It includes four convolutional blocks with 64, 128, 256, and 512 filters in the encoder. Each encoder block is followed by a batch-normalization layer, a ReLU[14] activation function, and a max-pooling layer. The bottleneck of the model has 1024 filters. The full U-Net architecture was initially pretrained on the BraTS brain tumor segmentation dataset. We used the U-Net encoder as the backbone of our model up to its final convolutional block, which contains 512 filters. But the max-pooling layer after this block was replaced with an average-pooling layer. The resulting architecture of the proposed model is shown in Figure 3.

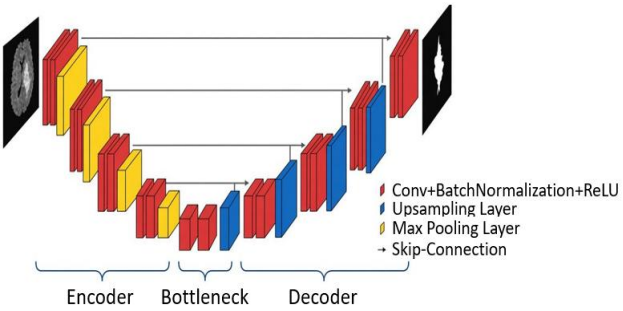

Figure 2. U-Net segmentation architecture used for pretraining on the BraTS dataset.

We implemented a prediction head for the U-Net encoder similar to that used with VGG-16. The resulting backbone and multi-task head are shown in Figure 3. This architecture was used in scenarios (c) and (d).

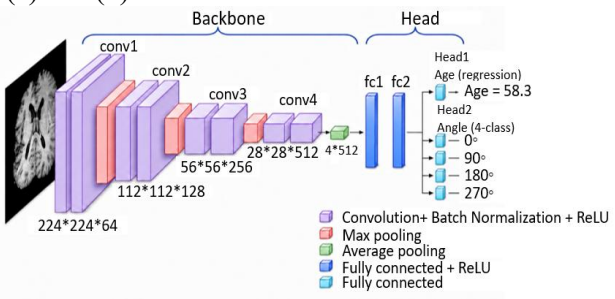

Figure 3. Structure of the proposed model with the U-Net encoder and multi-task head.

3. Pretraining

Using weights learned from one task to initialize another task is known as transfer learning[15]. This strategy enables models trained on related tasks to provide informative initial weights for the main brain age estimation task.
We employed two pretraining approaches—self-supervised DINO pretraining[16] and supervised tumor segmentation pretraining—to incorporate knowledge from auxiliary tasks.

3.1 Pretraining with DINO

We applied the self-supervised DINO framework to medical images [16]. DINO shares similarities with knowledge distillation and does not require labeled images. In knowledge distillation, a student network is trained to match the output of a teacher network, parameterized by θ_s and θ_t, respectively. As shown in Figure 4, given an input image x, both networks output probability distributions over K dimensions, denoted by P_s and P_t. The probability P is obtained by normalizing the output of each network with a softmax function. In the context of DINO, given an image x, we generate a set of V different views. This set includes two global views, x_1^g , x_2^g, and six local views at lower resolutions. All views are provided to the student, whereas only the global views are fed to the teacher. As a result, we minimize the loss as follows:

$$\min_{\theta_s} \sum_{x \in \{x_1^g, x_2^g\}} \sum_{x' \in V, x' \neq x} -P_t(x) \log P_s(x') \qquad (1)$$

Equation (1) illustrates the cross-entropy loss computed between the probability distributions of multiple views of an input image x, promoting the extraction of invariant features from distinct parts of the input image x[16].

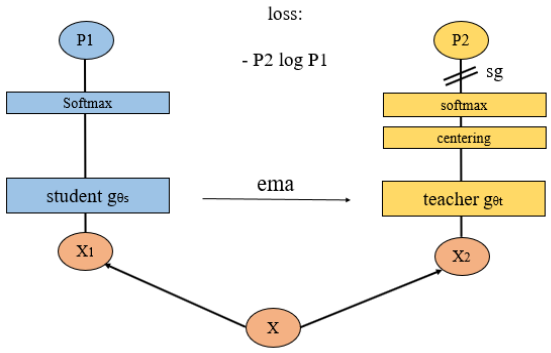

Figure 4. DINO architecture with teacher and student networks for an input image x [16].

To prepare global and local views of an input image x, we perform a center crop of 150 × 150 pixels. Since the views are selected randomly, they could have captured mostly black borders with little or no brain tissue, especially in local views. This initial crop reduces the likelihood of selecting such non-informative regions. Subsequently, we randomly select 25% to 100% of the center-cropped image for global views and 15% to 25% for local views. Both global and local views are subjected to random rotations of 0, 90, 180, or 270 degrees. Additionally, we apply a Gaussian smoothing filter to one of the global views and the same type of filter to all local views with

probabilities of 1 and 0.01, respectively. Following DINO's standard settings, we resize the global views to 224 × 224 pixels and local views to 96 × 96 pixels. Figure 5 illustrates the eight views generated from a sample MRI image.

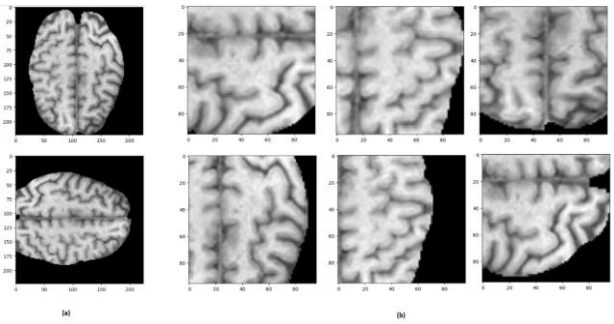

Figure 5. Examples of global and local views generated from a sample BraTS brain image: (a) two global views and (b) six local views.

Unlike conventional knowledge distillation, DINO does not update the teacher network through backpropagation. Instead, the teacher parameters are updated at each iteration using an exponential moving average (EMA) of the student parameters. Figure 6 illustrates the architecture of DINO, used during pretraining. The model employs a VGG-16 backbone and comprises student and teacher networks with identical architectures but distinct parameter sets.

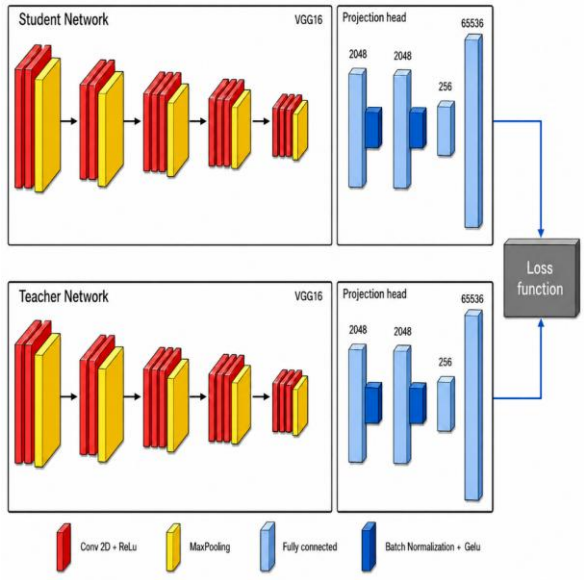

Figure 6. DINO architecture with a VGG-16 backbone.

Both student and teacher networks consist of a backbone (VGG-16) and a projection head. The projection head includes a three-layer perceptron with 2048, 2048, and 256 units, respectively. The first two layers are followed by a Gaussian error linear unit (GELU) activation and a batch normalization (BN) layer.

The three-layer perceptron is followed by an L2 normalization and a weight-normalized fully connected layer with K output dimensions, where K is set to 65,536 by default. Unless otherwise stated, we used the default DINO settings. The teacher-network backbone weights were subsequently used to initialize the downstream models.

3.2. Pretraining with Tumor Segmentation

For the second pretraining method, following the approach in[2], we pretrained the U-Net model for brain-tumor segmentation, using domain-specific knowledge to enhance the initialization of our network.

Tumor-segmentation pretraining is not intended to transfer knowledge of tumor biology to healthy brain aging. Instead, it provides the encoder with prior experience in identifying tissue boundaries, intensity transitions, and spatial relationships in brain MRI. These features may offer a better initialization than random weights for learning the distributed morphological changes associated with age. Because the segmentation and age-estimation tasks use T2-weighted BraTS and T1-weighted ADNI images, respectively, the transferred features are more likely to reflect anatomical structure than modality-specific intensity patterns.

We employed the Dice criterion[17], a commonly used loss function for tumor segmentation. This function measures the overlap between the estimated region ($\hat{Y}$) and the ground truth (Y), as shown in (2).

$$Dice(\hat{Y},Y)=\frac{2|\hat{Y}\bigcap Y|+s}{|\hat{Y}|+|Y|+s} \quad (2)$$

In (2), the smoothing parameter s was introduced for a more straightforward gradient calculation and was set to 1 during training.

To improve the effectiveness of tumor segmentation, we combined the Dice loss with binary cross-entropy (BCE) loss to promote more robust learning. BCE penalizes discrepancies between the predicted probabilities and the ground-truth labels. The Dice and BCE losses were summed to obtain the total training loss used for backpropagation. Equation (3) defines the resulting loss function.

$$TotalLoss=(1-Dice(\hat{Y},Y))+BCE(\hat{Y},Y) \quad (3)$$

After pretraining the U-Net for tumor segmentation, we removed the decoder and the final segmentation layer and retained only the encoder. The encoder was then connected to the shared fully connected layers and the two output

heads for age prediction and rotation classification. The resulting network was fine-tuned on the ADNI dataset. The lower layers of the encoder mainly capture basic image features, such as edges and local texture changes, while the deeper layers represent more complex spatial and anatomical information. During fine-tuning, these features were adjusted to identify the broader structural changes associated with brain aging rather than tumor localization.

4. Datasets

ADNI**:** In this work, we used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease[18].

The ADNI data were collected in several phases: ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. The dataset includes cases with Alzheimer's disease (AD), mild cognitive impairment (MCI), and cognitively normal (CN) status.

Similar to[2], we used MPRAGE T1-weighted MRI scans from all phases. We used images from cognitively normal (CN) participants in this dataset for our age estimation task.

The cognitively normal (CN) data from the ADNI-1 cohort were divided into validation and test sets according to the information provided on the LONI website. Additionally, the combined CN data from the ADNI-2 and ADNI-3 cohorts were used as the training set. Table 1 provides an overview of the demographic characteristics of each ADNI cohort.

Table 1. Demographic characteristics of the ADNI.

Dataset / use	samples	Sex (F/M)	Range	Mean ± SD
ADNI-1 / Validation	89	40/49	62–89.6	75.74 ± 5.16
ADNI-1 / Test	140	70/70	59.9–88.6	75.95 ± 4.94
ADNI-2 / Training	159	81/78	56–95	74.30 ± 6.87
ADNI-3 / Training	444	274/170	51–95	71.80 ± 8.36
Training total	603	355/248	51–95	72.45 ± 8.06

BraTS**:** The Brain Tumor Segmentation (BraTS) dataset contains structural brain MRI scans with expert annotations of tumor regions. Each subject has four standard MRI modalities: T1-weighted, T2-weighted, contrast-enhanced T1-weighted, and T2-FLAIR. The dataset images were preprocessed by registration to a standard anatomical atlas, resampling to an isotropic resolution of 1 mm, and removal of non-brain tissue through skull stripping.

For this study, we utilized BraTS 2020 to pretrain the U-Net model. The dataset comprises 369 volumes of individuals aged between 18 and 86 years, with an average age of 61.2 years[17].

5. Preprocessing

Images from the BraTS and ADNI-1 datasets are available in NIfTI format, whereas images from the ADNI-2 and ADNI-3 datasets are provided in raw DICOM format. We converted all DICOM images to NIfTI format using MRIcron software, version 2019. To ensure consistent orientation and spatial resolution, all ADNI images were registered to the MNI152 standard space and interpolated to a spatial resolution of 2 mm^3 using SPM12. Processing all ADNI images required approximately 30 hours. Skull stripping was then performed on the ADNI images using HD-BET [19] to remove the skull and other non-brain tissues (Figure 7).

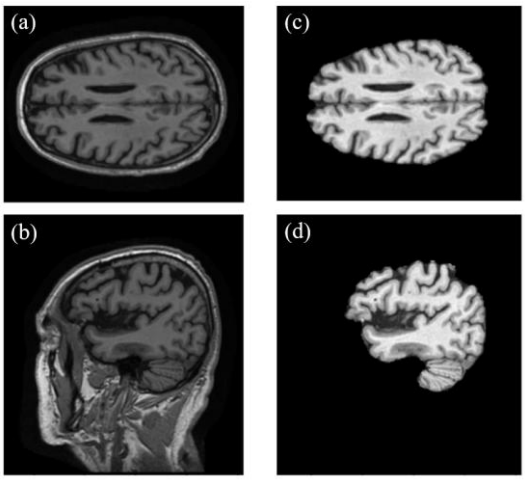

Figure 7. A sample of ADNI dataset images before and after skull removal in axial and sagittal views. MRI slices in (a) axial view and (b) sagittal view. The same MRI slices are shown after skull removal in (c) the axial view and (d) the sagittal view.

From each three-dimensional ADNI MRI scan, axial slices numbered 25 to 65 were extracted. The remaining slices, which contained little or no brain information, were discarded. This procedure yielded 40 two-dimensional axial slices from each three-dimensional MRI volume. This selection strategy is also consistent with the approach used in[2], which serves as the main comparison study for our work and similarly employed a predefined range of central axial slices. All images were resized to 224 × 224 pixels and normalized using a mean of 0.3125 and a standard deviation of 0.2419.

For the BraTS dataset, no preprocessing was performed other than slice selection and intensity normalization. For the VGG model used in the DINO pretraining task, axial slices numbered 36 to 116 were selected from each T1-weighted MRI volume, yielding 80 slices per volume. These data were normalized using the mean and standard deviation that we calculated from the BraTS dataset, which were 0.3804 and 0.3275, respectively. For pretraining the U-Net model for the segmentation task, six axial slices were selected from each T2-weighted MRI volume, beginning at slice 50 and using a step size of 10 slices. Each selected slice was normalized by dividing its voxel-intensity values by the maximum intensity value within that slice.

6. Training and Evaluation

Our brain age estimation models were developed based on prior studies reporting a strong association between biological brain age and chronological age in healthy individuals. Accordingly, the proposed models were trained and evaluated only on images from cognitively normal participants in the ADNI dataset. We adhered to the standardized split outlined in[2] to facilitate comparison. We split the preprocessed T1-weighted images from CN subjects in the ADNI-1 dataset to create our validation and test sets. The remaining data from the preprocessed ADNI-2 and ADNI-3 datasets constituted the training set. Each of the 40 input slices from a given volume was used to train the proposed models to predict brain age, using the subject's chronological age as the ground truth. In parallel, we implemented image rotation classification[20] as an auxiliary task to build our multi-task model. As described in Section 2.1, a four-class classification head was used. In addition to estimating the brain age of each slice, the network learned to predict the rotation class of each slice during training.

We chose rotation classification as the auxiliary task because predicting whether a brain image has been rotated requires the encoder to learn the overall orientation and spatial arrangement of brain structures. This encourages the model to use global anatomical information rather than relying only on local textures. Rotation itself does not provide direct information about age. Instead, it serves as an additional training signal that may help the shared encoder learn structural features useful for age estimation. For the rotation-classification task, each training slice was additionally rotated by 90°, 180°, and 270°, producing four versions of each image, including the unrotated image. The four class labels corresponded to rotation angles of 0°, 90°, 180°, and 270°.

The brain age head used mean squared error (MSE) loss, whereas the rotation-classification head used cross-entropy (CE) loss. The two losses were combined linearly to obtain the total loss used for backpropagation, as shown in (4). A weighting factor of 10 was applied because the CE loss had a smaller magnitude than the MSE loss.

$$Loss = MSE(\hat{y}_{age}, y_{age}) + 10 * CE(\hat{y}_{rot}, y_{rot}) \quad (4)$$

During validation, we evaluated the unrotated validation data using the mean absolute error (MAE), without considering either the rotation-classification head or the cross-entropy loss. For each MRI scan, the network predicted the age from each of the 40 slices. The 40 predictions were then averaged to obtain a subject-level age estimate, and the MAE was calculated between the predicted age and the true chronological age. The previously unseen test data were evaluated using the same procedure and the best-performing model weights. Preliminary experiments showed that excluding the cross-entropy loss from the validation criterion resulted in better performance than using the combined age-regression and rotation-classification losses in the validation phase.

7. Experiments and Results

7.1. Pretraining with DINO

We first initialized the model with single-channel ImageNet[21] weights using the Timm package instead of random initialization. The model was then trained on the preprocessed T1-weighted BraTS dataset, with a random data split of 80% for training and 20% for validation applied to both the teacher and student networks. The LARS optimizer[22] was used for 100 epochs with a batch size of 32 on the NVIDIA Tesla A100 GPU. The initial learning rate was set to 0.5 and was linearly warmed up during the first three epochs. After this warmup, the learning rate was decayed using a cosine scheduler[23] to 0.0048. The weight decay parameter started with a default value of 0.04 and was increased to the default value of 0.4 with the DINO cosine scheduler in each epoch.

Training was stopped at epoch 42 to mitigate overfitting. The best performance was obtained at epoch 35, with a training loss of 4.12 and a validation loss of 4.11. The loss curves are shown in Figure 8.

7.2 Pretraining with Tumor Segmentation

In addition to self-supervised DINO pretraining, we implemented supervised pretraining by training a U-Net model for brain-tumor segmentation.

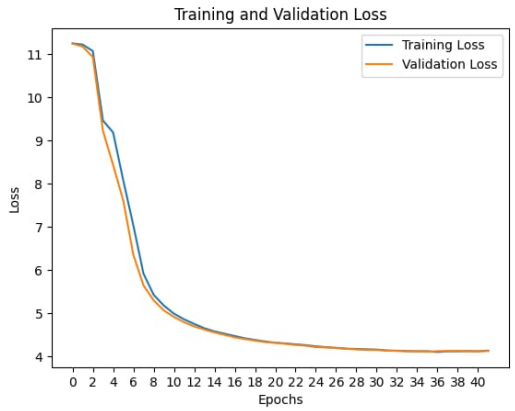

Figure 8. Training and validation loss curves for the DINO student network.

Two-dimensional axial slices from the preprocessed T2-weighted BraTS images were used to train the U-Net. We considered the ground truth of each slice in a binary manner for tumor and non-tumor areas. The data were randomly split into 85% for training and 15% for validation. The model was initialized with random weights and trained for 100 epochs using the Adam optimizer, a learning rate of 0.001, a batch size of 32, and an NVIDIA Tesla A100 GPU. No explicit early-stopping criterion, weight decay, or learning-rate scheduler was used. The learning rate remained constant, and the checkpoint with the best validation performance, obtained at epoch 82, was retained.

As described above, the total loss was defined as the sum of Dice loss and binary cross-entropy loss. Figure 9 shows the training and validation loss curves for the U-Net model. The best performance was obtained at epoch 82, with a training loss of 0.10, a training Dice score of 0.91, a validation loss of 0.17, and a validation Dice score of 0.86.

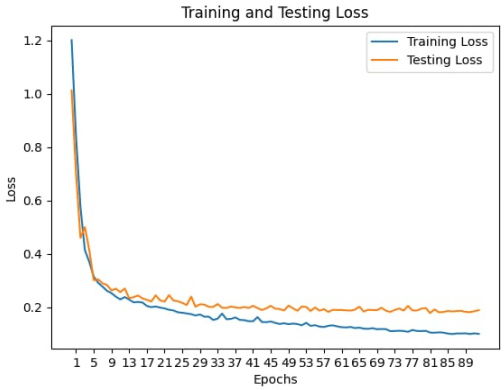

Figure 9. Training and validation loss curves for the U-Net model.

Figure 10 shows representative tumor segmentation outputs generated by the U-Net model.

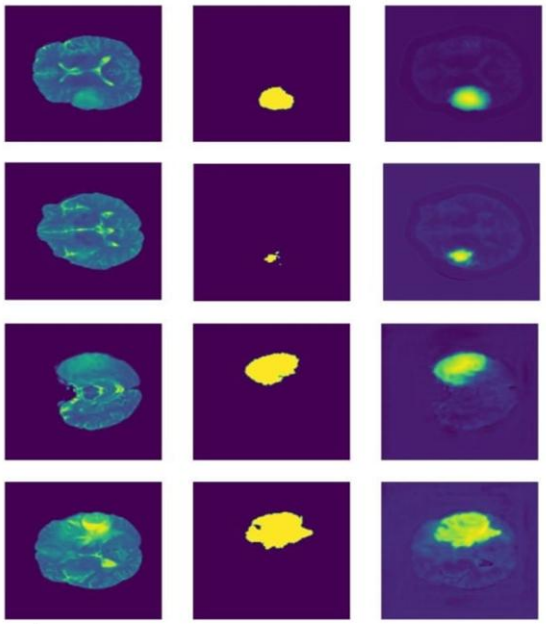

Figure 10. U-Net segmentation results. From left to right: input slice, ground-truth tumor mask, and predicted tumor mask.

7.3. Downstream Training and the Effect of Rotation

The multi-head network was initialized in three ways: with the DINO teacher-network weights, with U-Net encoder weights obtained from tumor-segmentation pretraining, or with random weights. We also evaluated a single-head version of the architecture that estimated only brain age and did not include the rotation head or rotation loss. This network was initialized once with the DINO teacher network weights and once with random weights, and its performance was compared with that of the multi-head network, providing valuable insights into the benefits of the multi-task architecture. For the DINO-pretrained models, the backbone parameters were frozen during the first three epochs. For the other pretrained models, all weights remained trainable from the start because this setting produced better empirical performance. All models were trained using the Adam optimizer for 50 epochs, with a batch size of 64 and a gradient-clipping value of 3. Weight decay was set to 0.04 and gradually increased to 0.4 using a cosine scheduler. All models without pretrained weights were trained with a learning rate of 25e-6, whereas all models with pretrained weights were trained with a learning rate of 3e-6 on the ADNI dataset images. Each model was trained three times on NVIDIA Tesla GPUs to ensure reliable results. The average results of three runs are summarized in Table 2. To determine whether a multi-task model with rotation as an auxiliary task or a single-task model with rotation as data augmentation is

more effective, we conducted tests using the single-task model with rotation data augmentation and the VGG-16 backbone for brain age estimation.

Table 2. Mean performance over three runs for the evaluated models. Italicized values indicate the best performance in each column. Performance is reported as mean absolute error (MAE) on the validation and test sets.

Head	Backbone	Pretrainig weights	Validation MAE	Test MAE
Multi-Task	U-Net	Random	4.31 ± 0.414	4.48 ± 0.390
Multi-Task	U-Net	Tumor segmentation	3.68 ± 0.048	3.56 ± 0.104
Single-Task	VGG-16	Random	3.90 ± 0.042	3.72 ± 0.057
Multi-Task	VGG-16	Random	3.81 ± 0.021	3.58 ± 0.005
Single-Task	VGG-16	DINO (teacher)	3.70 ± 0.016	3.36 ± 0.025
Multi-Task	VGG-16	DINO (teacher)	3.54 ± 0.106	3.27 ± 0.089

We used the same hyperparameters, random initialization, and training, validation, and test splits as the corresponding single-head baseline. Each training image was augmented by rotations of 90°, 180°, and 270°. The experiment was repeated three times, and Table 3 reports the mean performance across the runs.

Table 3. Comparison of single-task models with and without rotation-based data augmentation and the multi-task model using rotation classification as an auxiliary task. Values are averaged over three runs on the ADNI dataset.

Head	Pretraining weights	Rotation data augmentation	Validation MAE	Test MAE
Single-Task	Random	YES	3.94 ± 0.018	3.70 ± 0.059
Single-Task	Random	NO	3.90 ± 0.042	3.72 ± 0.057
Multi-Task	Random	NO	3.81 ± 0.021	3.58 ± 0.005

The two single-task models in Table 3 achieved similar performance, and rotation-based data augmentation produced only a small improvement in test MAE. In contrast, the multi-task model that used rotation classification as an auxiliary task achieved a lower test MAE than either single-task model. Under this experimental setting, rotation classification was therefore more effective as an auxiliary task than rotation used solely for data augmentation.

8. Discussion and Conclusion

Overall, both supervised tumor segmentation pretraining and self-supervised DINO pretraining improved performance relative to their corresponding non-pretrained models. The benefit of tumor segmentation pretraining was demonstrated by comparing scenarios (c) and (d), whereas the contribution of DINO pretraining was assessed through the comparisons of scenarios (a) versus (e) and scenarios (b) versus (f), respectively. The best result was obtained with a DINO-pretrained VGG-16 backbone and a multi-task head for brain age regression and rotation classification, which achieved a test MAE of 3.27 years. This result suggests that DINO pretraining and auxiliary rotation prediction provide complementary benefits by helping the model learn more informative age-related features from brain MRI. From a scientific perspective, the findings show that improved brain age estimation does not necessarily require introducing a new backbone architecture; it can also be achieved through more effective pretraining and auxiliary-task learning. The results also indicate that structural representations learned from unlabeled brain images can be transferred and refined for age regression. From a practical perspective, neither DINO pretraining nor rotation classification requires additional manual annotations, which is particularly useful when labeled neuroimaging data are limited. Across the evaluated VGG-16 configurations, the multi-task models outperformed their corresponding single-task counterparts, suggesting positive synergy rather than negative interference between age estimation and rotation classification. Specifically, comparing scenarios (a) and (b), adding the rotation-classification task reduced the test MAE of the non-pretrained VGG-16 model from 3.72 to 3.58 years. Similarly, comparing scenarios (e) and (f), the test MAE of the DINO-pretrained VGG-16 model decreased from 3.36 to 3.27 years. These improvements suggest that the auxiliary rotation task helps the shared backbone learn richer and more generalizable features for brain age estimation. However, this interpretation is based on prediction performance, as no direct gradient-conflict analysis was conducted. We used MAE as the main evaluation metric because it reports the prediction error in years and makes our results easy to compare with previous brain age studies. However, a single MAE value cannot fully describe model performance, as it does not show the direction of the prediction errors or how accuracy varies across different age groups. This is an important limitation in our study because the

ADNI sample mainly includes older adults, and the number of participants is not evenly distributed across the age range. As a result, the reported MAE may be more representative of the age groups with more samples, while larger errors in less represented groups may remain unclear. Furthermore, since the model was developed using cognitively normal participants from ADNI, its performance may not directly generalize to younger individuals, clinical populations, or MRI data collected at other institutions. Differences in demographic characteristics and imaging conditions, including sex, ethnicity, education, recruitment site, scanner type, and acquisition protocol, may also affect the model's predictions. Future studies should therefore examine performance within separate age groups, include additional evaluation metrics, and validate the model on independent and more diverse datasets. The current implementation should therefore be regarded as a research prototype rather than a clinically ready system. Clinical deployment would require external and multicenter validation, evaluation across different scanners and acquisition protocols, and prospective clinical assessment. In addition, the relatively high computational and memory requirements of the VGG-16 backbone may limit its practical use in routine clinical settings. Future work should investigate lighter architectures with lower computational costs.

Table 4 presents a contextual comparison between the proposed method and previous brain-age estimation studies that used MRI data from the ADNI dataset. However, these results cannot be compared directly in all cases. Many earlier studies did not clearly report which ADNI phases, participants, or scans were assigned to the training, validation, and test sets. In addition, differences in cohort composition, age distribution, preprocessing steps, model inputs, and evaluation procedures make it difficult to interpret the reported MAE values as a strict ranking of the methods. Pacheco et al. [2] identified the lack of a common and reproducible data-splitting protocol as an important limitation in previous brain-age studies. They noted that, even when a public dataset such as ADNI was used, the exact samples included in each experimental subset were often not specified. This makes it difficult for other researchers to reproduce the experiments or perform a fair comparison. To address this issue, they reported their data-selection and splitting procedure in detail and used clearly defined ADNI subsets. We followed the same approach in the present study and adopted their data-splitting protocol as closely as possible.

Among the studies listed in Table 4, the work of Pacheco et al. provides the most relevant comparison with our study. Both studies used the same ADNI cohort split and estimated subject-level brain age by averaging the predictions obtained from 40 axial slices of each MRI volume. Our proposed model, which combines a DINO-pretrained VGG-16 backbone with a multi-task head for age regression and rotation classification, achieved a test MAE of 3.27 years. Although this value is numerically lower than the other results reported in Table 4, it should not be interpreted as evidence of absolute superiority. Comparisons with studies other than Pacheco et al. should be made cautiously because of the differences in data selection and experimental design.

Table 4. Comparison of training, validation, and test data in selected brain-age estimation studies.

Study	Training data	Validation data	Test data	Test MAE (years)
More et al. [24]	2953 healthy participants aged 18–88 years from CamCAN, IXI, eNKI, and 1000BRAINS	Model selection and cross-dataset evaluation within the four training datasets; no fixed validation set reported	209 CN participants from ADNI	6.56
Lam et al. [25]	Training folds from 631 CN ADNI participants using subject-level 10-fold cross-validation	Fixed validation set including 56 CN ADNI participants	Held-out CN folds; one scan retained per participant for testing	3.96
Ly et al. [26]	757 healthy amyloid-negative participants from ADNI, IXI, and OASIS-3	Not separately reported	51 CN amyloid-negative participants from ADNI	3.70
Poloni et al. [27]	1189 CN participants from ADNI and IXI	Not separately reported	151 CN ADNI participants older than 70 years	3.66
Pacheco et al. [2]	606 CN participants from ADNI-2 and ADNI-3	89 CN participants from ADNI-1	140 CN participants from ADNI-1	3.284
Proposed model	603 CN participants from ADNI-2 and ADNI-3	89 CN participants from ADNI-1	140 CN participants from ADNI-1	3.27

References

[1] I. Hwang, E. K. Yeon, J. Y. Lee, R. Yoo, K. M. Kang, T. J. Yun, S. H. Choi, C. Sohn, H. Kim and J. Kim, "Prediction of brain age from routine T2-weighted spin-echo brain magnetic resonance images with a deep convolutional neural network," *Neurobiology of Aging*, vol. 105, p. 78-85, 2021.

[2] B. M. Pacheco, V. .. de Oliveira, A. B. Antunes, S. D. Pedro and D. Silva, "Does pre-training on brain-related tasks results in better deep-learning-based brain age biomarkers?," *In: Naldi, M.C., Bianchi, R.A.C. (eds) Intelligent Systems. BRACIS 2023. Lecture Notes in Computer Science()*, vol 14196. Springer, Cham. https://doi.org/10.1007/978-3-031-45389-2_13.

[3] J. H. Cole, S. J. Ritchie, M. E. Bastin, M. C. Valdés Hernández, S. Muñoz Maniega, N. Royle, J. Corley, A. Pattie, S. E. Harris, Q. Zhang, N. R. Wray, P. Redmond, R. E. Marioni, J. M. Starr, S. R. Cox, J. M. Wardlaw, D. J. Sharp and I. J. Deary, "Brain age predicts mortality," *Molecular Psychiatry*, vol. 23, p. 1385–1392, 2018.

[4] J. H. Cole, R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector and G. Montana, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker," *NeuroImage*, vol. 163, p. 115-124, 2017.

[5] B. A. Jonsson, G. Bjornsdottir, T. E. Thorgeirsson, L. M. Ellingsen, G. Bragi Walters, D. F. Gudbjartsson, H. Stefansson, K. Stefansson and M. O. Ulfarsson, "Brain age prediction using deep learning uncovers associated sequence variants," *Nature communications*, vol. 10.1, no. 5409, 2019.

[6] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, 2016.

[7] V. M. Bashyam, G. Erus, J. Doshi, M. Habes, I. Nasrallah, M. Truelove-Hill, D. Srinivasan, L. Mamourian, R. Pomponio, Y. Fan, L. J. Launer, C. L. Masters, P. Maruff, C. Zhuo, H. Völzke, S. C. Johnson, J. Fripp and N. Koutsouleris, "MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14468 individuals worldwide," *a journal of neurology*, vol. 143, no. 7, p. 2312–2324, 2021.

[8] H. Peng, W. Gong, C. F. Beckmann, A. Vedaldi and S. M. Smith, "Accurate brain age prediction with lightweight deep neural networks," *Medical Image Analysis*, vol. 68, no. 101871, 2021.

[9] C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman and R. Collins, "UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age," *PLoS Med*, vol. 12, no. 3, 2015.

[10] D. M. Hedderich, A. Menegaux, B. Schmitz-Koep, R. Nuttall, J. Zimmermann, S. C. Schneider, J. G. Bäuml, M. Daamen, H. Boecker, M. Wilke, C. Zimmer, D. Wolke, P. Bartmann, C. Sorg and C. Gaser, "Increased Brain Age Gap Estimate (BrainAGE) in Young Adults After Premature Birth," *Front Aging Neurosci*, vol. 13, no. 653365, 2021.

[11] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351, p. 234-241, 2015.

[12] Z. Zhang, Q. Liu and Y. Wang, "Road Extraction by Deep Residual U-Net," in *IEEE Geoscience and Remote Sensing Letters*, vol. 15, p. 749-753, 2018.

[13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[14] A. F. Agarap, "Deep Learning using Rectified Linear Units (ReLU)," arXiv preprint arXiv:1803.08375, 2018.

[15] M. Raghu, C. Zhang, J. Kleinberg and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems, NY, USA*, 2019.

[16] C. Mathilde, T. Hugo, M. Ishan, J. Hervé, M. Julien, B. Piotr and J. Armand, "Emerging Properties in Self-Supervised Vision Transformers," in *IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada*, 2021.

[17] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom and R. Wiest, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, p. 1993-2024, 2015.

[18] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski and M. W. Weiner, "Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization.," *Neurology*, vol. 73, no. 3, p. 201-209, 2010.

[19] F. Isensee, M. Schell, I. Pflueger, G. Brugnara, D. Bonekamp, U. Neuberger, A. Wick, H. Schlemmer, S. Heiland, W. Wick, M. Bendszus, K. Maier-Hein and P. Kickingereder, "Automated brain extraction of multisequence MRI using artificial neural networks," *Human Brain Mapping*, vol. 40, no. 17, p. 4952–4964, 2019.

[20] S. Gidaris, P. Singh and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *ICLR*, 2018.

[21] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA*, 2009.

[22] Y. You, I. Gitman and B. Ginsburg, "Large batch training of convolutional networks," arXive preprint: 1708.03888, 2017.

[23] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," in *International Conference on Learning Representations (ICLR)*, 2016.

[24] S. More, G. Antonopoulos, F. Hoffstaedter, J. Caspers, S. B. Eickhoff and K. R. Patil, "Brain-age prediction: A systematic comparison of machine learning workflows," *NeuroImage*, vol. 270, 2023.

[25] P. Lam, A. H. Zhu, I. B. Gari, N. Jahanshad and P. M. Thompson, "3d grid-attention networks for interpretable age and alzheimer's disease prediction from structural mri," 2020.

[26] M. Ly, G. Z. Yu and H. T. Karim, "Improving brain age prediction models: incorporation of amyloid status in Alzheimer's disease," *Neurobiology of Aging*, vol. 87, p. 44-48, 2020.

[27] K. M. Poloni and R. J. Ferrari, "A deep ensemble hippocampal CNN model for brain age estimation applied to Alzheimer's diagnosis," *Expert Systems with Applications*, vol. 195, no. 116622, 2022.

تخمین سن بیولوژیکی مغز با استفاده از خودنظارتی چندوظیفه‌ای و شبکه‌های عصبی عمیق از پیش آموزش‌دیده

زهراسادات سجادی، سهیل حمزه‌بیگی و محسن سریانی*

دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.

ارسال ۲۰۲۶/۰۴/۲۰؛ بازنگری ۲۰۲۶/۰۶/۰۵؛ پذیرش ۲۰۲۶/۰۸/۰۴

چکیده:

با افزایش سن جمعیت جهان، روش‌های قابل اعتماد برای ارزیابی سلامت مغز و تغییرات مرتبط با سن به طور فزاینده‌ای اهمیت پیدا می‌کنند. سن مغز یک نشانگر زیستی امیدوارکننده برای سلامت مغز است و روش‌های یادگیری ماشینی امکان تخمین آن را از داده‌های تصویربرداری عصبی فراهم کرده‌اند. با این حال، برای تخمین دقیق سن مغز، به استراتژی‌های آموزشی مؤثر نیاز است. این مطالعه یک چارچوب چندوظیفه‌ای مبتنی بر شبکه عصبی کانولوشن دوبعدی (2DCNN) برای تخمین سن مغز از اسکن‌های تصویربرداری تشدید مغناطیسی (MRI) در مجموعه داده ADNI پیشنهاد می‌دهد. این چارچوب از ستون فقرات کدگذار VGG-16 و U-Net با سرهای جداگانه برای رگرسیون سن و طبقه‌بندی چرخش تصویر استفاده می‌کند. ما اثرات چندین استراتژی پیش‌آموزش، از جمله پیش‌آموزش خودنظارتی با استفاده از چارچوب DINO و پیش‌آموزش تحت نظارت از طریق قطعه‌بندی تومور مغزی را ارزیابی کردیم. بهترین پیکربندی، متشکل از یک ستون فقرات VGG-16 که پیش آموزش‌دیده با DINO است و یک سر پیش‌بینی چندوظیفه‌ای، به میانگین خطای مطلق (MAE) 3.27 سال دست یافت که با روش‌های گزارش‌شده قبلی قابل رقابت است. نتایج نشان می‌دهد که ترکیب یادگیری انتقالی با یادگیری چندوظیفه‌ای، عملکرد را نسبت به مدل‌های تک‌وظیفه‌ای مربوطه بهبود می‌بخشد، که نشان می‌دهد این ترکیب از یادگیری نمایش‌های غنی‌تر و تعمیم‌پذیرتر ویژگی‌ها پشتیبانی می‌کند.

کلمات کلیدی: یادگیری عمیق، DINO، یادگیری چندوظیفه‌ای، پیش‌آموزش، یادگیری انتقالی.