IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 November 2024, accepted 21 December 2024, date of publication 6 January 2025, date of current version 13 January 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3526520*

RESEARCH ARTICLE

# Brain Age Prediction Using a Lightweight Convolutional Neural Network

**FATMA ELTASHANI[1], MARIO PARRENO-CENTENO[2], JAMES H. COLE[3], JOÃO PAULO PAPA[4], (Senior Member, IEEE), AND FUMIE COSTEN[1], (Senior Member, IEEE)**

[1]Electrical and Electronic Engineering Department, The University of Manchester, M13 9PL Manchester, U.K.
[2]Comprehensive Cancer Centre, Guy's Hospital, SE1 3SS London, U.K.
[3]Centre for Medical Image Computing and Dementia Research Centre, University College London, WC1V 6LJ London, U.K.
[4]School of Sciences, São Paulo State University, São Paulo 05508-090, Brazil

Corresponding author: Fatma Eltashani (fatma.eltashani@manchester.ac.uk)

**ABSTRACT** Much interest has recently been drawn to brain age prediction due to the significant development in machine learning and image processing techniques. Studies based on brain magnetic resonance images showed a strong relationship between the brain ageing process and accelerated brain atrophy, suggesting using brain age prediction models for early diagnosis of neurodegenerative disorders, such as Parkinson's, Schizophrenia, and Alzheimer's disease. However, data availability, acquisition protocols diversity and models' computational complexity remain limiting factors for clinical adoption. This study proposes a low-complexity convolutional neural network (CNN) model that tackles these challenges, focusing on three main aspects: performance accuracy, computational complexity, and adaptability to new, external datasets. We developed a brain-age prediction system using a minimally preprocessed T1-weighted MRI images with a multi-site dataset of healthy individuals covering the whole human lifespan (2251 subjects, age range 6-90 years). We proposed a lighter version of the Simple Fully Convolutional Network (SFCN) that contain only 1.2 million parameters. Computational load was further reduced by cropping the brain images. Finally, we employed transfer learning approach to achieve domain adaptation to external, unseen sites. We demonstrated that leveraging the cropped brain images reduced the computational time for training by 50%, maintaining a comparable accuracy to using the entire brain. The model achieved a Mean Absolute Error (MAE) of 3.557 for the full brain and 4.139 for the cropped images with a Pearson correlation $r = 0.988$ between the full and cropped brain predictions when evaluated on the same test set. Domain adaptation of our model to new external data showed a significant improvement in the prediction performance, reducing MAE from 7.219 to 4.750 for full brain images and from 12.107 to 5.770 for the cropped images. This study is the first to demonstrate comparable prediction accuracy using only a small segment of a 3D full brain MRI scan. Our results show that it is feasible to build lightweight CNN models trained on small-scale, heterogeneous datasets and fine-tuned to new external clinical data, making significant steps toward practical clinical application.

**INDEX TERMS** Biological age estimation, brain imaging, brain ageing, convolutional neural network, deep learning, magnetic resonance imaging.

## I. INTRODUCTION

Brain ageing produces a functional decline and a probability of developing neurodegenerative illnesses such as Alzheimer's disease (AD) [1]. Although the brain ageing process generally involves autophagy to prevent brain atrophy, the efficiency of autophagy differs significantly from one individual to another. This variation factor may contribute to the different ages at which people develop age-related diseases. These relate to lifestyle, health condition, and environment [2]. Human chronological age is defined as the age of a person from the date of their birth, while

The associate editor coordinating the review of this manuscript and approving it for publication was Marco Giannelli.

the biological age is determined by the effect of neural development events involving progressive and regressive processes in the brain [3].

A reliable measure of the difference between chronological and biological age could strongly indicate future health risks and diseases. The possibility of problems related to cognitive ageing and diseases could be identified by how much a person's brain diverges from average healthy brain ageing trajectories. Age estimation has been approached using various media and methods in the past, each leveraging different types of data and techniques such as facial images [4] and DNA methylation [5]. However, brain age prediction based on Magnetic Resonance Images (MRI) has recently become the most popular method due to the limitations of non-imaging approaches as they tend to indicate the whole body rather than the ageing of individual organs, which is not always accurate [6].

With the considerable development of machine learning (ML) and deep learning, many studies have been carried out on brain datasets using structural or functional data. Despite the various methods, many challenges remain for more reliable and robust brain age prediction systems. Shallow learning techniques such as Support Vector Machines (SVM), Relevance Vector Regression (RVR), and Gaussian Process Regression (GPR) have been heavily implemented in the neuroscience age prediction field [7], [8], [9]. However, despite their good performance, the necessity of performing feature extraction cannot be avoided.

Feature extraction processes can be time-consuming and require deep domain knowledge. Additionally, they can introduce subjectivity and may sometimes affect the generalisation ability of models. The model performance significantly depends on the quality of the manually extracted features [10]. Such problems are overcome by how neural networks and deep learning methods approach statistical modelling in neuroimaging. The advancement in computer infrastructures has allowed the use of sufficiently large imaging data in their raw, minimally preprocessed form. Convolutional Neural Networks (CNNs) achieved high performance for computer vision tasks such as classification and segmentation. They are widely utilised to predict brain age based on MR images either using two- [11], [12] or three-dimensional [13], [14] CNN backbones. Though more computationally expensive, the latter can incorporate information between MRI slices.

Despite the success of these methods, clinical adoption faces several significant challenges. Various acquisition protocols, scanner types, and demographic differences in the training data are crucial factors that impact model performance. Current state-of-the-art frameworks often rely on large datasets that are either not publicly available or encompass a narrow age range, failing to cover the entire lifespan. Another critical challenge is the computational complexity and training time of these models; training on large MRI datasets can take considerable time, sometimes extending to days or even weeks. While methods such as dimensionality reduction and optimised architectures can improve efficiency, they may also compromise accuracy. Therefore, it is essential to find a balance between efficiency and accuracy. Furthermore, distributional differences due to variabilities in acquisition protocols and image resolution limit the clinical applicability of ML models. Models trained on one dataset may not perform well when applied to the same task in a different environment, especially when the distribution of the new external test data differs from that of the training data [15]. Therefore, there is a need for high-performance, reliable models trained with easily obtainable, representative data that can perform well across different domains.

This study aims to provide a robust age prediction system that addresses the aforementioned challenges limiting the clinical application of current algorithms: (1) Develop a low-complexity CNN model trained on publicly available datasets acquired from various sources, covering almost the entire human lifespan. (2) Utilise MR image size reduction to enhance training time efficiency without compromising performance accuracy. (3) Investigate the adaptability of the model to data from external, unseen sites by employing a transfer learning approach.

The remainder of this work is organised as follows: Sections II, III and IV present the dataset, proposed approach, and experimental results respectively, and Section V discusses the outcomes and states final remarks.

## II. MATERIALS AND DATASET

### A. COMPREHENSIVE MULTISITE DATASET

The neuroimaging dataset collected in this study comprises T1-weighted MRI scans of 2251 healthy subjects (male/female = 1346/ 905, mean age = 32.96 ± 24.29, age range 6-90 years), stored in the NIFTI file format. These images were compiled from various publicly accessible neuroimaging databases and were acquired at 1.5T or 3T using various types of scanner. All subjects included in this study were free of major neurological or psychiatric diagnoses. Hence, we assume that the chronological age of the participants approximates their biological age, as no clinical standard for biological age is available. Table 1 summarises the collected data, including the age and sex of the participants. Due to data cleaning, the number of samples per dataset is slightly less than in the original sources. Images with missing information or inconsistent labeling were excluded from the study. Fig. 1 shows the age range distribution for each data site.

### B. DATA AGE IMBALANCE

The data collected in this study is multi-site, resulting in a significant imbalance in the distribution of participants' ages and the number of MR images obtained from each site. As illustrated in Fig.2, the majority of the samples are concentrated in the younger age groups, with almost half of the subjects lying in age ranges below 30 years.

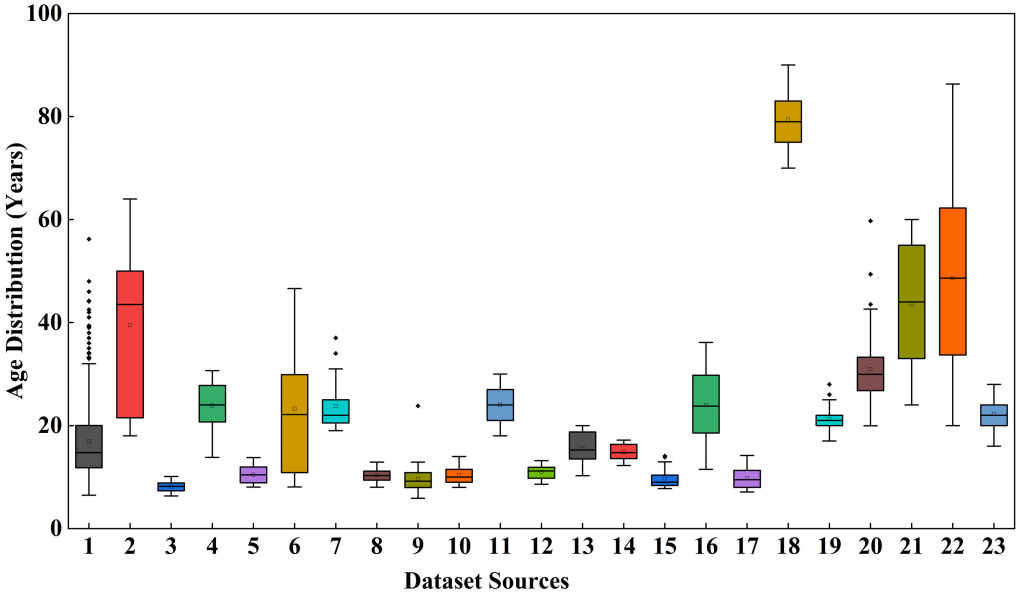


**FIGURE 1.** Age range distribution of MRI samples collected from various dataset sources: source number 1 to 23 represent the dataset ABIDE, ABIDEII (BNI1), ABIDEII (EMC1), ABIDEII (ETH1), ABIDEII (GU1), ABIDEII (IP1), ABIDEII (IU1), ABIDEII (KKI1), ABIDEII (NYU1), ABIDEII (OHSU1), ABIDEII (ONRC2), ABIDEII (SU2), ABIDEII (TCD1), ABIDEII (MIA1), ABIDEII (UCD1), ABIDEII (UCLA1), ABIDEII (USM1), BeigingEn, Berlin, Clevend, IXI, ADNI and Train, respectively.

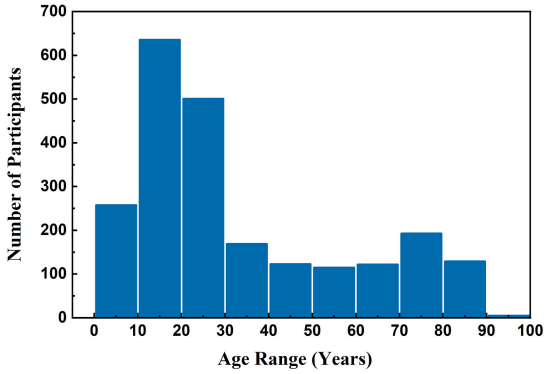


**FIGURE 2.** Number of data in each age range. The dataset covers the whole human lifespan with an average age of 30 years.

**TABLE 1.** Dataset information representing the number of samples, age range, average age, and and male/female distribution.

| Dataset | Participants | Age range | Mean | Sex (M/F) |
|---|---|---|---|---|
| Berlin | 49 | 20 – 60 | 31.00 | 24/25 |
| Cleveland | 31 | 24 – 60 | 43.55 | 11/20 |
| Train | 39 | 16 – 28 | 22.26 | 11/28 |
| Beijing En. | 180 | 17 – 28 | 21.22 | 73/107 |
| ABIDE | 554 | 6 – 56 | 16.91 | 458/96 |
| ABIDE II | 550 | 6 – 64 | 33.77 | 364/186 |
| ADNI | 270 | 70 – 90 | 79.38 | 132/138 |
| IXI | 562 | 20 – 86 | 48.64 | 250/312 |
| **Total** | **2251** | **6 – 90** | **32.96** | **1346/905** |

This substantial imbalance in data distribution poses a considerable challenge for predictive modeling, particularly in regression models. An imbalanced dataset can easily lead to misleading performance metrics as MAE and $R^2$ may be overall acceptable but can hide poor performance on underrepresented age groups. Additionally, the variability in image sources, acquisition protocols, and scanner settings across sites may result in a poorly generalised model.

To address the data imbalance issue, stratification techniques were applied when partitioning the data into training, validation, and testing sets. This ensures that each subset accurately represents the original distribution, resulting in a robust model that maintains high accuracy and reliability throughout the entire spectrum of ages.

### C. IMAGE PREPROCESSING

All acquired scans are preprocessed using the FMRIB Software Library (FSL) [16]. The importance of preprocessing is to have a common geometrically standardised coordinate system so that the model learns the anatomical pattern of the brain scans rather than focusing on other features, such as non-brain tissues. First, a brain extraction is carried out to remove all non-brain tissues. Then, an image registration step aligns each brain image to the standard shape and pose of the MNI152 average brain image. The MNI152 is a template based on an average of 152 MRIs of healthy individuals developed by the Montreal Neurological Institute [17]. Previous research showed that grey matter (GM) is a strong indicator of brain ageing [18]. Unlike the white matter, the GM volume declines linearly and constantly throughout the ageing process [19]. Therefore, all 3D scans in this study are segmented to extract their GM content. The resultant GM scans are then padded or cropped to have a standard size of $91 \times 109 \times 91$ voxels.

## III. METHODOLOGY AND APPROACH

In this study, we address the primary challenges encountered by brain age prediction algorithms, focusing on three main aspects: performance accuracy, computational complexity, and adaptability to new external datasets. We begin by assessing the performance of our algorithm using full-sized grey matter brain images. This initial evaluation serves to establish a baseline against which we can measure the impact of subsequent modifications. To enhance computational efficiency, we then modify the algorithm's input to process only regions of the brain images which include the most informative regions by reducing the image size. This step is designed to retain critical information while minimising computational load. Finally, we validate the performance of our models in both the full-sized and reduced-image versions—on an external dataset. This is to test the adaptability of our approach, ensuring its applicability in real-world clinical settings.

### A. PROPOSED CNN MODEL

A simple and low-complexity 3D Convolutional Neural Network is developed for age prediction. The architecture of the model is inspired by the Simple Fully Convolutional Neural Network (SFCN), a deep CNN that won the Predictive Analysis Challenge for brain age prediction, proposed in [14] and [20]. Unlike the original SFCN, which treats the task as a soft classification problem where the age prediction is defined as the average of all predictions, our model is tailored for regression, directly predicting the subject's age as a continuous output. Furthermore, to enhance the model's computational efficiency, the number of convolution blocks is reduced to 4 as opposed to the original structure, which compromises 5 blocks. This resulted in reducing the number of parameters from 3 million [14] to only 1.2 million, which leads to faster training times and lower memory usage. Additionally, in our architecture, the batch normalisation layer is placed after the activation function rather than before it to enhance the training stability and model performance. This decision is based on studies suggesting that post-activation normalisation can lead to more stable training dynamics and potentially better generalisation [21], [22]. By normalising the output of the activation function, we aim to maintain desirable statistical properties of the network activations, thereby improving convergence during training. Fig. 3 illustrates the proposed model, consisting of four convolutional blocks and additional fully-connected layers.

Blocks 1 to 4 include 3D convolutional layers with kernel sizes of $3 \times 3 \times 3$, a $2 \times 2 \times 2$ max-pooling layers followed by a ReLU activation function and a batch normalisation layer. The primary rationale behind these blocks is to decrease the spatial dimensions of the input, hence, reducing the number of trainable parameters. This approach significantly improves the efficiency and memory usage. By the end of the fourth block, the dimensionality of the input is significantly reduced. The following block (block 5) consists of a smaller kernel size ($1 \times 1 \times 1$) to reduce each feature map without affecting the spatial dimensions while increasing non-linearity. Further, block 6 includes a global average pooling layer for feature map reduction and a dropout layer for generalisation purposes, followed by a $1 \times 1 \times 1$ convolution layer, acting as a dense layer to further compact the learned features. The feature map is then further reduced using sequential fully-connected layers with 128 and 64 neurons, respectively. A final dense layer with a linear activation function generates the output. The output is then flattened to produce a single continuous numerical output representing the predicted age. This structure ensures an efficient processing of volumetric data, optimizing the network for the quick and effective prediction of age from MRI scans.

### B. BRAIN AGE ESTIMATION

We first evaluate the performance of the architecture using full-sized brain grey matter images. The dataset, described in Section II, is divided into training, validation and testing sets with splits of 80%, 10%, and 10%, respectively, each representing the age distribution that is identical to the original dataset in Fig.2. To ensure robustness and reliable performance metrics, we apply a repeated random subsplitting strategy, where the data is randomly split into training, validation, and test sets 5 times. The final results reported are the average performance across these five independent runs. Models are trained from scratch using Adaptive Moment Estimation Optimiser (ADAM) [23]. To fairly compare the training times of different models, we opted to train each model for a fixed number of epochs. This ensures consistency in training times, as convergence criteria, such as early stopping, can result in variable training times, complicating the comparative analysis. Different numbers of epochs were evaluated and 100 epochs were selected because they provided a good balance between performance and training time. The best weights of the validation set are saved for the accuracy calculation of the test set. A batch size of 16 and a learning rate of $10^{-3}$ are used throughout the training process. The loss function for these tasks is Mean Absolute Error (MAE). These parameters are selected based on hyperparameter tuning to find the best performing combination. All experiments are run on a NVIDIA A100-SXM4-40GB GPU and developed using Python 3.10.12.

### C. MR IMAGE SIZE REDUCTION

Although changes in brain structures are heterogeneous, making it challenging to define which regions of the brain are responsible for the ageing process, the cerebellum volume represents about 10% of the human brain, suggesting the high computational power of this structure [24]. Studies have reported that the correlation between cerebellum volume and age ranges from $-0.30$ to $-0.43$ [25], [26]. Additionally, research studies have indicated that substantial annual declines appear in the hippocampus and amygdala [27], [28].

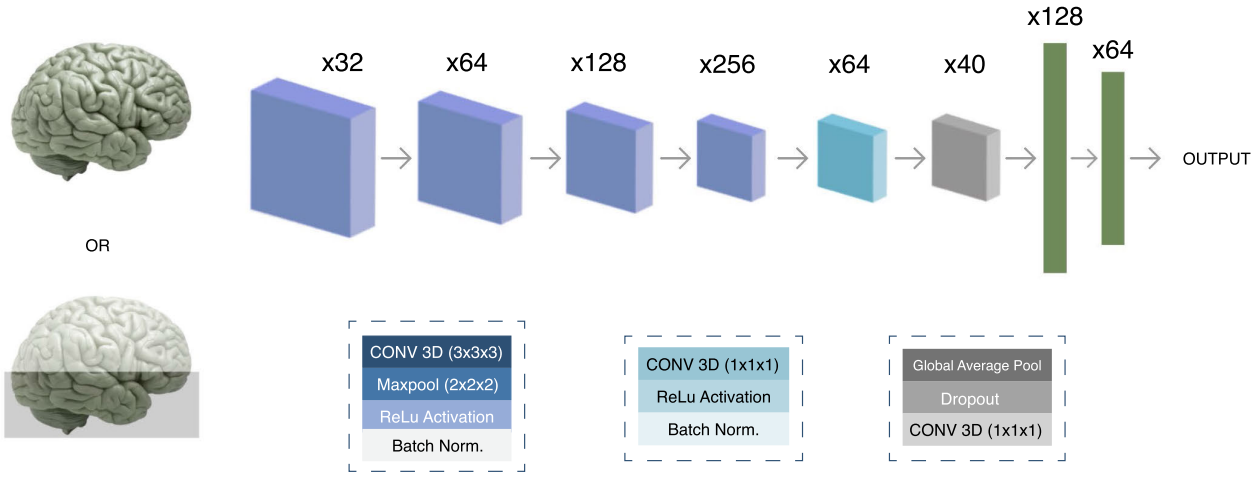


**FIGURE 3.** Overview of the proposed 3D Convolutional Neural Network (CNN) architecture for age prediction. The input to the network is either the full brain image with grey matter extracted or a cropped version (lower part of the brain). Blocks 1 to 4 are highlighted in purple and consist of 3D convolutional layers (3 × 3 × 3 kernels), max-pooling layers (2 × 2 × 2), ReLU activation, and batch normalization, as depicted in the color-coded block underneath. The number of output channels is indicated above each block. Block 5 is shown in light blue, consisting of 1 × 1 × 1 convolutional layer to reduce feature map dimensions, followed by batch normalisation and ReLU activation. Block 6, in light grey, includes a global average pooling layer, dropout for regularisation, and a final 1 × 1 × 1 convolutional layer. This block is followed by fully connected layers, producing the final continuous output for age prediction.

The hippocampus, in particular, exhibits a yearly atrophy rate ranging from 0.79% to 2.0% [29]. Therefore, we define the Region of Interest (ROI) as the brain region that includes the cerebellum, most of the amygdala, and the hippocampus. We crop the brain image into a smaller segment, encompassing major brain structures and optimizing size.

Fig.4 shows the sagittal, coronal, and axial cross sections of the brain, which pass through the cerebellum, amygdala, and hippocampus. These structures are located mainly within the range $0 \leq z \leq 35$ of the entire brain, where $z$ is the vertical index variable. Thus, we set the ROI as $0 \leq z \leq 35$,the remaining two dimensions of $x$ and $y$ kept the same as the dimensions of the original image. The resultant image $91 \times 109 \times 35$ image includes our ROI, which is approximately one third of the entire brain data. This method of slicing does not affect the coordinate system of the data, and therefore no further registration is required.

### D. EVALUATION ON NEW EXTERNAL DATASET

The proposed algorithms, using full-sized and cropped brain images, are evaluated on an unseen external dataset consisting of the Dallas Lifespan Brain Study (DLBS) and the ABIDE II (SDSU) dataset. DLBS is recognised for its focus on studying the early stages of a healthy brain's progression towards Alzheimer's disease. The DLBS dataset includes 314 individuals spanning the adult lifespan with a balanced distribution between different age groups (male/female: 117/197; mean age: 55 years; age range: 21-89 years).

Although the DLBS dataset includes a sufficient number of participants to evaluate the model, it lacks participants from younger age groups, unlike the original dataset, which includes almost the entire lifespan. Therefore, we include the SDSU subset of the ABIDE II dataset. This subset is not included in the original training set, and is chosen due to its age range, which resembles the missing range in comparison to the original dataset. It consists of images from 25 young individuals (age range: 8-18 years, mean age: 13 years, male/female: 23/2). This selection aims to closely resemble the original dataset and provide a realistic representation of the actual human lifespan, thus enabling a robust assessment of the model's ability to generalise across varied human conditions. This resulted in a total external dataset consisting of 339 images (male/female: 140/199; mean age: 51.66 ± 24.29 years; age range: 8-89 years).

**TABLE 2.** External unseen dataset information representing the number of samples, age range, average age, and male/female distribution.

| Dataset | Participants | Age range | Mean | Sex (M/F) |
|---|---|---|---|---|
| DLBS | 314 | $21-89$ | 55 | 117/197 |
| ABIDE II (SDSU1) | 25 | $8-18$ | 13 | 23/2 |
| **Total** | **339** | $\mathbf{8-89}$ | **51.66** | **140/199** |

#### 1) WEIGHTED LOSS FUNCTION

As discussed in Section II, the acquired data set is significantly unbalanced, with a higher concentration of samples in the younger age group of adults. To effectively address this imbalance, a weighted mean absolute error (MAE) loss function is implemented to assign higher importance to samples from underrepresented age groups, particularly older individuals. We hypothesise that this approach improves

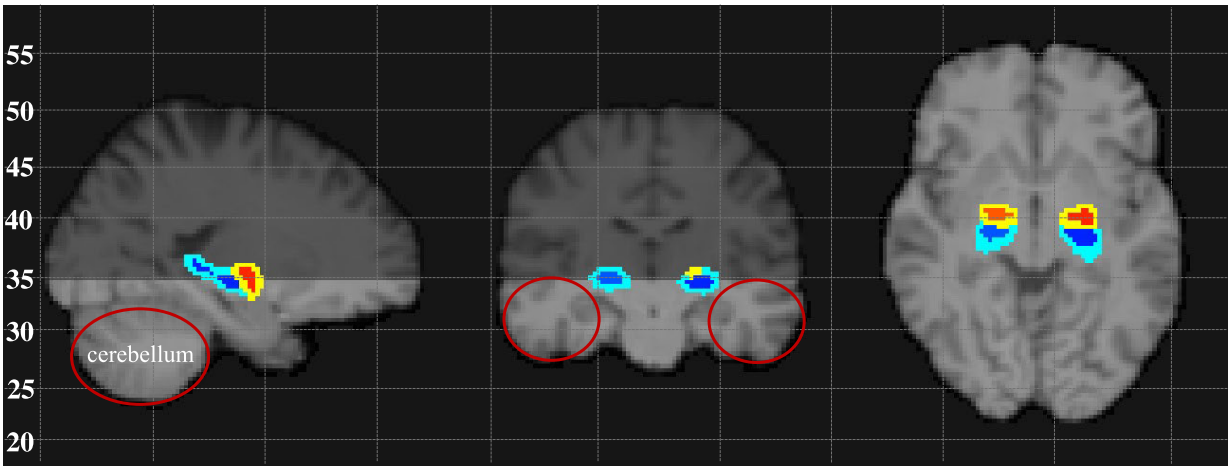


**FIGURE 4.** An example of MRI data used in the brain age prediction models representing the sagittal (left), coronal (middle), and axial (right) planes of the image after the minimal preprocessing and normalisation to the MNI152 space using FSL software. The hippocampus is highlighted in blue, while the amygdala is highlighted in red. The cerebellum is circled in red. The numbers on the left represent the values of the $z$-axis used for image cropping; these values are not to scale. Cropping was performed at $z = 35$ to include the cerebellum and the majority of the hippocampus and amygdala.

generalisation across all age groups and prevents the model from being biased towards the majority age groups. Furthermore, it ensures that older individuals, which are often more clinically significant, are accurately predicted.

In this scenario, we assign a weight to each sample, for $N$ number of samples, that is proportional to the true age $y_i$. This weight is then multiplied by the MAE calculated based on the prediction $\hat{y}_i$. Therefore, the weighted loss $L_w$ can be expressed as:

$$L_w = \frac{1}{N}\sum_{i=1}^{N}\left(|y_i - \hat{y}_i|\right)\cdot\left(\frac{y_i}{100.0}\right) \tag{1}$$

### 2) TRANSFER LEARNING

from scratch.

As we will demonstrate in Section IV, the model's performance on the external dataset exhibits a discrepancy due to domain shift caused by the imbalanced age distribution in the training data. This issue is effectively addressed by fine-tuning the pre-trained model on new external subjects. Fine-tuning, a widely used transfer learning technique [15], [30], [31], adapts a pre-trained model to a new task or domain with minimal additional training. This approach offers simplicity and efficiency compared to training models from scratch, making it particularly suitable for clinical applications, where variability in resolution and field strengths across MRI scans poses significant challenges.

Consider a model trained on a source domain $\mathcal{D}_S$. In transfer learning, we use the pre-trained model $M_S$ as a starting point for the new target domain $\mathcal{D}_T$. Let $\theta_S$ be the weights learned from the source domain. The pre-trained model will be initialised with $\theta_S$ and fine-tuned to update the weights $\theta_T$ for the target domain $\mathcal{D}_T$ using a smaller learning rate. This process ensures that the model adapts to the new target domain efficiently, leveraging the knowledge gained from the source domain.

In this study, we evaluate baseline models trained using either MAE loss or weighted loss on full brain or cropped brain images. The evaluation on an unseen external test set is conducted through the following approaches: (1) External Assessment without Fine-Tuning: We evaluate the model performance on the external dataset without any additional fine-tuning. (2) External Assessment with Fine-Tuning: We apply transfer learning by splitting the external dataset into 80% training, 10% validation, and 10% testing sets. (3) Comparison with Age-Matched Data Distribution: We compare the performance of the models between the external dataset used as it is and when its age distribution is matched to the original training set. All results are averaged across the performance of 5 models, each trained using 5-fold cross-validation in the baseline experiments.

For the fine-tuning process, we initially freeze the first five convolution blocks, allowing only the last block and fully-connected layers to adapt to the external dataset $\mathcal{D}_T$. The model is trained with a learning rate of $10^{-4}$ and employs checkpoints to save the best-performing model on the validation set. Subsequently, the frozen layers are unfrozen to fine-tune the model further, learning the detailed features of the new dataset with a reduced learning rate of $10^{-5}$. The rest of the hyperparameters and the choice of optimiser remain identical to those used in the baseline model trained from scratch.

## IV. EXPERIMENTAL RESULTS

### A. SIZE REDUCTION IMPROVES EFFICIENCY

The predictive accuracy for both full and cropped brain images using our proposed CNN algorithm is presented in Table 3. All models were trained on a training set

($n = 1801$), a validation set ($n = 225$), and tested on a test set ($n = 225$). The age distribution was stratified to ensure an accurate representation of all age groups across the training, validation, and testing sets. Table 3 showcases the results of evaluating five models while varying the combination of the training, validation and testing sets. All the models are trained from scratch and evaluated on the test datasets. The performance measures include Mean Absolute Error (MAE) in years, coefficient of determination $R^2$, Pearson $r$, and training time in minutes. A lower MAE indicates that the model's predictions are more accurate on average, which is important in the context of brain age prediction as it implies greater reliability in assessing age-related changes in brain structure. A high $R^2$, close to 1 suggests that the model effectively captures the age-related patterns in brain structure, indicating a robust and meaningful prediction across a broad population. Similarly, a higher Pearson $r$ implies a stronger correlation between the actual and the predicted age, highlighting the model's reliability. Using full brain grey matter images, the model achieved an accurate brain age prediction with a mean MAE of 3.557 years and a Pearson $r$ of 0.978. In contrast, using only the lower segment of the brain yielded an MAE of 4.139 years and a Pearson $r$ of 0.971, with the training time being only half of that required for the entire brain image. The consistency across experiments, as indicated by the low standard deviations, suggests that the model performs reliably across different subsets of the data. We further evaluated the performance of the full brain and cropped brain models when trained on the same data split and tested on the same group of participants and the results showed a Pearson correlation $r = 0.988$ between the two predictions. The strong correlation proves the stability of the model based on different inputs demonstrating highly consistent predictions.

### B. DATA IMBALANCE EFFECT

In both the full brain and cropped brain experiments, the brain age gap is calculated and plotted in Fig. 5. This is to visualise the difference of the prediction error as a function of the chronological age of the test sets. The effect of the age group on the model's prediction is represented by the slope of the line. Although the accuracy of the cropped image of the brain is slightly lower when compared to using the full brain image, we can see that the slope of the age gap line fit is almost identical, suggesting that using the cropped brain images does not affect specific age groups more than others compared to using the full brain images. Another important consideration is the prediction error around the mean. Brain age estimation models generally suffer from overestimation for younger subjects and underestimations for older subjects, where the prediction error is close to zero around the mean of the dataset. Although the overestimation error for subjects younger than the mean is considered small, the underestimation for older subjects could have a significant impact when the model is applied to an external dataset. In this study, we explore if this issue is caused by the bias introduced by the imbalanced training set by considering: (1) Age-matching the distribution of the external dataset to ensure that any overestimation or underestimation is identified; (2) Modify the learning algorithm using the weighted loss function discussed in Section III-D1 and evaluating them with the external dataset; (3) Employing transfer learning methods to fine-tune the trained model on the external dataset before evaluation.

### C. ADAPTATION TO EXTERNAL DATASET

External validation is crucial for assessing the performance and generalisability of our predictive model. In this section, we use a completely independent dataset that was not involved in the training process. Details of the dataset are described in Section III-D.

In this study, the distribution of the external dataset differs from that of the training set, a phenomenon known as domain shift. We hypothesise that this leads to a drop in model performance due to certain age groups being inadequately represented in the original training set.

#### 1) WEIGHTED LOSS FUNCTION

The performance accuracy of our predictive model is detailed in Table 4. Here, we evaluate the accuracy of models trained using the weighted loss function described in Section III-D1, compared to models trained using the original MAE loss. The reported results represent the average of five cross-validation models. The models were tested on the internal dataset (n = 225), an external dataset with the original distribution (n = 339), and the age-matched distribution (n = 195) without fine-tuning, as well as age-matched test set of (n = 30) with fine-tuning.

For the internal assessment, we compared the performance of the model using the MAE loss, as illustrated in Section IV-A, to the performance using the weighted loss, which gives more importance to underrepresented samples. The performance metrics of the models trained with the weighted loss function yielded the following results: The full brain model achieved a mean MAE of 4.276, a mean of $R^2$ of 0.945, and a mean Pearson $r$ of 0.974. In contrast, the cropped brain model achieved a mean MAE of 4.804, a mean of $R^2$ of 0.926, and a mean Pearson $r$ of 0.964. Compared to the standard MAE loss, we find that there is no statistically significant difference between the two loss functions ($P = 0.187$ for the full brain and $P = 0.123$ for the cropped brain). Although the standard MAE loss provided slightly better performance for both input types than the weighted loss, the disparity between the full brain and cropped image performances was very similar. Fig.7 shows a visual representation of the performance accuracies using standard MAE and weighted loss functions across the different adaptation methods for full and cropped brain images.

For the external dataset, we further assessed the model's performance using the original distribution against an age-matched distribution to minimise the effect of biased models.

**TABLE 3.** Predictive accuracy comparison of the model's performance on the test set using the brain MRI images versus using the cropped brain MRI together with the models' training time in minutes. The results obtained by training 5 different models and evaluating their performance on the test sets, the overall accuracy is calculated by taking the mean and standard deviation (STD) of the five models. Performance metrics presented as MAE mean absolute error, R2 coefficient of determination and Pearson r.

| Model | Full brain image | | | | Cropped brain image | | | |
|---|---|---|---|---|---|---|---|---|
| | MAE | $R^2$ | Pearson r | Training time (mins) | MAE | $R^2$ | Pearson r | Training time (mins) |
| 1 | 3.602 | 0.953 | 0.977 | 6.74 | 4.054 | 0.944 | 0.973 | 3.53 |
| 2 | 3.453 | 0.959 | 0.980 | 6.67 | 4.347 | 0.940 | 0.973 | 3.46 |
| 3 | 3.756 | 0.945 | 0.977 | 6.80 | 4.237 | 0.928 | 0.966 | 3.54 |
| 4 | 3.664 | 0.953 | 0.978 | 6.74 | 3.857 | 0.949 | 0.975 | 3.52 |
| 5 | 3.317 | 0.960 | 0.980 | 6.76 | 4.200 | 0.939 | 0.970 | 3.56 |
| **Mean** | **3.557** | **0.954** | **0.978** | **6.74** | **4.139** | **0.940** | **0.971** | **3.52** |
| **STD** | **0.189** | **0.008** | **0.003** | **0.05** | **0.174** | **0.005** | **0.002** | **0.03** |

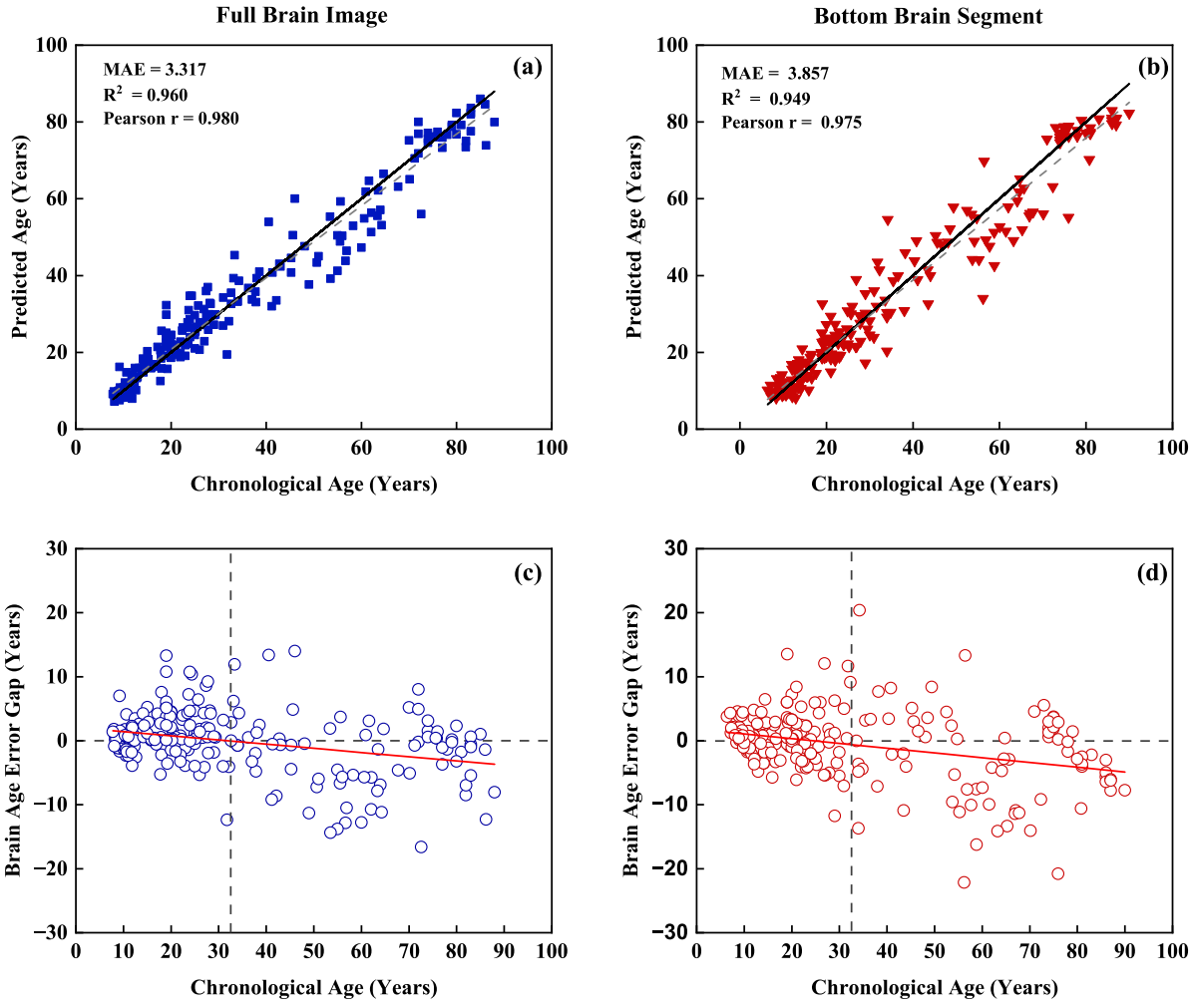


**FIGURE 5.** Comparison between the best-performing brain age prediction algorithms performance when the input is the entire brain image and when the input is the cropped brain image. Scatter plots show the predicted age versus chronological age with full brain images (a, dark blue), cropped brain images (b, dark red). In both plots, the dashed grey lines represent the linear regression and the black line indicates the ideal prediction. Scatter plots (c and d) represent the brain age gap versus the chronological age for both full brain (blue circles) and cropped brain images (red circles). The red lines represent the linear regression lines, and the horizontal dashed grey lines represent ideal prediction where the error is zero. The vertical dashed grey lines show the prediction around the mean age of our dataset. MAE, mean absolute error, $R^2$ coefficient of determination.

As expected, the performance of the model improved when the distribution was matched to the original training set. Additionally, the results suggest that there was no significant difference between the performance measures of the two loss functions when the full brain image was used. However, the weighted loss function outperformed the original MAE loss when the cropped brain image was used, with a mean MAE of 9.930, a mean $R^2$ of 0.712, and a mean Pearson $r$ of 0.937, compared to an MAE of 12.107, $R^2$ of 0.569, and a Pearson $r$ of 0.913 for the MAE loss function. While matching the

**TABLE 4.** Predictive accuracy of the model's evaluation on the internal and external test sets. The results were obtained following the application of the MAE loss function, weighted loss function, original distribution of internal and external datasets and age-matched distribution of external datasets. Performance metrics were calculated based on the models generated by the five experiments and presented as mean values. MAE mean absolute error, $R^2$ coefficient of determination and Pearson r. The standard deviation of the reported results are visualised in Fig.7.

| Method | Loss function | Dataset | Full brain image | | | Cropped brain image | | |
|---|---|---|---|---|---|---|---|---|
| | | | MAE | $R^2$ | Pearson r | MAE | $R^2$ | Pearson r |
| Internal assessment | MAE Loss | original | **3.557** | **0.954** | **0.978** | **4.139** | **0.940** | **0.971** |
| | Weighted loss | original | **4.276** | **0.945** | **0.974** | **4.804** | **0.926** | **0.964** |
| External assessment without fine-tuning | MAE Loss | external | 8.608 | 0.760 | 0.949 | 14.413 | 0.351 | 0.876 |
| | | age-matched | 7.219 | 0.848 | 0.967 | 12.107 | 0.569 | 0.913 |
| | Weighted loss | external | 8.279 | 0.780 | 0.946 | 11.754 | 0.559 | 0.909 |
| | | age-matched | 7.131 | 0.854 | 0.963 | 9.930 | 0.712 | 0.937 |
| External assessment with fine-tuning | MAE Loss | external | 6.160 | 0.886 | 0.953 | 6.625 | 0.848 | 0.931 |
| | | age-matched | **4.750** | **0.919** | **0.968** | **5.770** | **0.903** | **0.954** |
| | Weighted loss | external | 6.029 | 0.890 | 0.954 | 6.149 | 0.883 | 0.947 |
| | | age-matched | 4.669 | 0.921 | 0.964 | 6.464 | 0.873 | 0.941 |

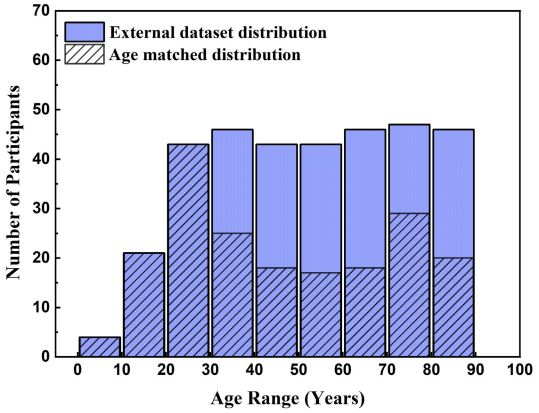


**FIGURE 6.** Representing the original age distribution of the external dataset and the age-matched distribution to the dataset used to pre-train the models shown in Fig.2. Age groups from 10 to 20 years were not matched due to the small number of samples from ABIDEII (SDSU).

age distribution improved the performance in this controlled experimental setting, this option might not be available in real-world applications where the data are highly variable and may not match the training set. Therefore, opting for the weighted loss function in these scenarios would be better, as demonstrated in Fig.7.

### 2) TRANSFER LEARNING

Baseline models trained using both MAE loss and custom weighted loss exhibited a significant reduction in generalisability when evaluated in the external dataset without additional fine-tuning. This discrepancy is likely due to the varying data acquisition protocols between different sites. Notably, the impact of this variability was more pronounced when using the cropped brain compared to the full brain image. Although custom weighted loss demonstrated better performance in the external dataset than the MAE loss (MAE = 9.930, Pearson $r = 0.937$), the deviation from the baseline internal assessment remained substantial (Internal assessment: MAE = 4.804, Pearson $r = 0.964$).

The application of transfer learning to fine-tune the baseline models markedly improved their performance on the external dataset, resulting in accuracy metrics closely aligning with those from the internal assessment when the MAE loss is used with the age-matched data distribution, with $P = 0.176$ (larger than 0.05) indicating that there is no statically significant difference between the internal and the external evaluation. Similarly to the initial evaluation without fine-tuning, there was no significant difference in performance between the MAE loss and the custom weighted loss for the full brain images (MAE loss: MAE = 4.75, Pearson $r = 0.968$ ; Weighted loss: MAE = 4.669, Pearson $r = 0.964$). However, for the cropped brain image, the MAE loss outperformed the weighted loss (MAE loss: MAE = 5.770, Pearson $r = 0.954$ ; Weighted loss: MAE = 6.464, Pearson $r = 0.941$).

Interestingly, for the cropped brain models, the custom weighted loss performed better on the external data without age-matching than with age-matching. This suggests that the weighted loss function may be more robust to certain types of variability in external datasets, making it a suitable choice when the distribution of external data differs from the original dataset.

## V. DISCUSSION

This paper proposes an estimation of brain chronological age for healthy individuals based on minimally preprocessed T1-weighted MR images using a low-complexity convolutional neural network. Our focus is to address the main challenges that limit the clinical adoption of brain age prediction systems, including achieving a balance between performance accuracy, computational efficiency, and adaptability to unseen domains and datasets. We demonstrated that an accurate prediction of brain age can be achieved using a small-scale publicly available dataset representative of the entire lifespan. Furthermore, we showed that computational efficiency can be significantly improved by using only the cropped brain, which reduces training time by half. Finally, we illustrated that models pre-trained on small-scale but diverse datasets can generalise to new external datasets with some fine-tuning.

Due to the complex nature of 3D CNN-based brain age estimation systems, it is challenging to provide a fair

comparison of our model's performance to the literature. The huge variation in datasets, age ranges, model architectures, and computational powers are limiting factors. Therefore, this work was compared to previous studies based on the brain images used, the training time of CNN models, and the dataset size.

### A. BRAIN AGE ESTIMATION

As a direct comparison to Peng et al. [14], our 3D CNN architecture was inspired by the SFCN model proposed in their work. However, we introduced several key modifications to enhance performance and stability. Both architectures include repeated blocks of convolution + batch normalisation + max-pooling and activation layers. This block architecture was repeated four times in our work, whereas it was repeated five times in the original SFCN architecture. Moreover, we found that positioning the batch normalisation layer after the activation layer yielded to a more stable training compared to the original SFCN architecture. Furthermore, the SFCN treats age prediction as a soft classification problem where each possible age is a separate class. In contrast, our model is a regression problem that directly predicts the age.

Performance-wise, when comparing our model to Peng et al. [14], it is important to consider the characteristics of the datasets used. Our model's full brain achieved an MAE of 3.55 years on a dataset of 2251 publicly available images, whereas Peng et al. reported an MAE of 2.19 years using the UK Biobank dataset with 12949 images. The superior performance of UK Biobank is expected due to the large enough representation of variability in data. In contrast, although our dataset is small, it contains images from diverse sources and is acquired using different protocols and scanner types. This diversity increases the variability within the data, thus increasing the complexity of the training process.

However, our model demonstrates better performance compared to the SFCN when trained from scratch on the multi-site PAC 2019 GM dataset [20], as evaluated by Peng et al. [14]. In their study, the SFCN achieved a mean absolute error (MAE) of 3.93 years on 2,198 images from PAC 2019. Although the PAC 2019 dataset is similar in size to ours, it includes only adult subjects aged 17 to 90, whereas our dataset spans a broader age range of 6 to 90 years, representing the full human lifespan. Furthermore, a recent study by Li et al. [32] employed a larger dataset with 3,479 samples and reported a slightly lower MAE of 2.941 years. However, this dataset also covers only adult subjects within the 18- to 90-year age range. The inclusion of a wide age span in our study, from childhood through to old age, presents additional complexity to the prediction task due to the substantial variations in brain morphology across different developmental and atrophic phases. This performance comparison suggests that our model is robust, as MAE values tend to be lower in studies focusing on narrower age ranges. Additionally, error values are typically smaller when the predictions are closer to the mean age of the dataset [33].

Therefore, it is challenging to achieve low MAE values with a wide age, particularly with a limited sample size. Studies that include age ranges similar to ours, such as [34], [35], [36], and [37] used significantly large datasets (minimum 8379 scans). The performance of our model highlights its robustness and training quality. Fig.8 shows a comparison of the performance of our model with the literature, where the axes $y$ and $x-$ represent the MAE, and the age range of the datasets (maximum age $-$ minimum age), respectively. The area of the circles is proportional to the number of samples in each dataset of these studies. The result of our study is highlighted in red. It can be seen that the MAE values tend to be smaller for studies that include a large number of samples with narrower age ranges, and the MAE values tend to be higher for studies that included a smaller number of samples. Therefore, achieving a low MAE with a small number of samples and a wide age range is an indication of a well-trained model.

### B. TRAINING TIME EFFICIENCY

Without significantly compromising performance, our network was designed with only 1.2 million parameters, significantly fewer than most state-of-the-art models such as VGG16 (138 million parameters) and ResNet-50 (25 million parameters) [43], [44]. The SFCN model, which inspired our architecture, contains approximately 3 million parameters. Although alternative large models make use of complexity reduction tools such as pruning and distillation, these techniques introduce an additional post-processing stage, increasing the complexity of the pipeline. In contrast, our lightweight CNN model ensures complexity reduction by design, without the need for such tools, achieving efficiency directly during training.

CNNs are particularly well-suited for brain age prediction due to their ability to efficiently capture spatial hierarchies and local dependencies in MRI data. Unlike transformers or MLP backbones, which often require larger datasets and computational resources to achieve comparable performance on spatially dependent tasks, CNNs leverage inductive biases such as locality and translational invariance. These characteristics make CNNs computationally efficient and effective at modeling the changes associated with brain ageing.

Furthermore, utilising the lower brain segment reduced the number of data points per sample, reducing training time by 50% while maintaining comparable accuracy. This approach effectively reduced memory usage and computational load, making our model feasible for deployment in clinical settings. Although some studies reduce complexity by defining brain atlases or using predefined ROIs, our method relies on cropping based on standard brain dimensions. This eliminates

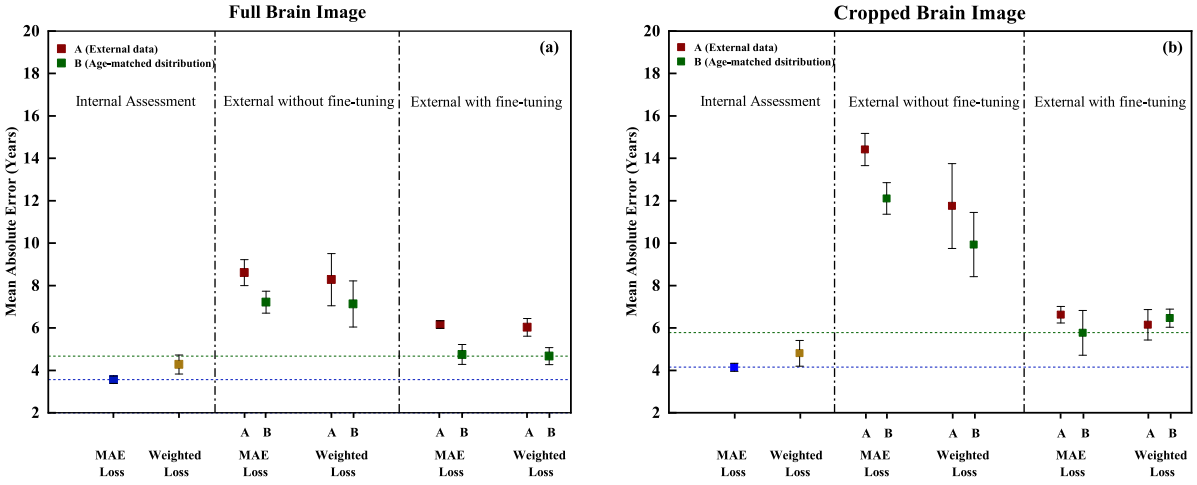


**FIGURE 7.** Error plots demonstrating the performance accuracy of the trained model on the internal and external datasets with an without fine-tuning. The error plots show the average MAE and the standard deviation of the five trained models for full brain images (a) and cropped brain images (b). The results were obtained by applying MAE loss function, weighted loss function, original distribution of the internal and external datasets and age-matched distribution of external dataset (blue, yellow, red and green respectively). The blue dashed line shows the lowest MAE achieved on the internal dataset, and the green dashed line represents the lowest MAE value achieved on the new external dataset.

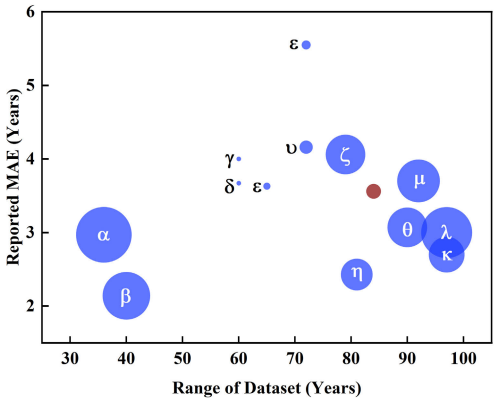


**FIGURE 8.** A comparison of the proposed model's performance (highlighted in red) against some state-of-the-art works that used CNNs and MR images of healthy individuals, $\alpha$ [38], $\beta$ [14], $\gamma$ [11], $\delta$ [39], $\xi$ [40], $\upsilon$ [13], $\epsilon$ [30], $\zeta$ [41], $\eta$ [42], $\theta$ [35], $\mu$ [34], $\lambda$ [14], $\kappa$ [37]. The area of the circles is proportional to the number of samples used in each study.

the need for atlas-based preprocessing or anatomical priors, resulting in a model that is more adaptable to diverse datasets and does not require domain-specific anatomical knowledge of brain structures.

Few studies have reported the time required to train models for brain age prediction. For example, Cole et al. [13] reported 18 to 48 hours to train their CNN model, and Peng et al. [14] mentioned that it took 50 hours to train the SFCN on the UK Biobank dataset. Li et al. [32] reported training times of 36, 110, and 237 minutes for their 3D-DenseNet, 3D-ResNeXt, and 3D-Inception-v4 models, respectively, using four NVIDIA GeForce RTX 3090 GPUs. In comparison, our model achieved average training times of 6.743 minutes for the full brain and 3.519 minutes for the cropped brain, using an NVIDIA A100-SXM4-40GB GPU.

To the best of our knowledge, this is the first study to achieve comparable prediction accuracy (within 0.7 years of full-brain predictions) using only a small segment of a 3D brain MRI scan. This significant improvement in computational efficiency reduces training time, operational costs, and memory usage, enabling faster real-time feedback and model updates, which are critical for real-world clinical applications.

### C. ADAPTAION TO EXTERNAL DATASETS

The substantial variability in MRI data acquisition—including differences in resolution, field strength, and orientations—poses significant challenges for developing models that generalise well across diverse datasets. An essential element of this study is addressing this issue by demonstrating the effectiveness of transfer learning in achieving promising results in overcoming these challenges.

Transfer learning involves leveraging knowledge gained from one task to enhance performance on a related but distinct task. Although this technique has been used in brain age prediction studies, many of these studies rely on models pre-trained on unrelated tasks, such as image recognition using datasets like ImageNet [30], [34].

In contrast, our research adopts a novel approach by initially training a model from scratch using a small-scale, diverse dataset with a broad age range and utilising both full and lower segments of brain MRI images. We then apply transfer learning to fine-tune this model on specific data from external, unseen sites. This methodology offers two significant benefits: (1) Task-specific pre-training: Models pre-trained on the same task can leverage specific features and patterns directly relevant to the task, facilitating better

adaptation and improved performance. (2) Efficacy of small-scale Datasets: Models pre-trained on small-scale but diverse datasets can achieve accurate predictions when fine-tuned with smaller datasets. This finding challenges the common notion that models must be pre-trained on large-scale datasets to achieve high performance.

Therefore, our study demonstrates that transfer learning can be effectively applied to models pre-trained on the same task. Pre-training on small-scale but diverse datasets can yield robust models capable of fine-tuning for specific tasks with high accuracy. This approach underscores the adaptability and efficiency of using targeted, representative datasets in model training and fine-tuning for brain age prediction.

## VI. LIMITATIONS AND CONCLUSION

This study addresses the challenges facing brain age prediction systems; however, several limitations remain to be considered. Despite the satisfactory performance of the CNN model, understanding the learned model remains challenging due to the black-box nature of deep learning. Furthermore, while this study aims to enhance our understanding of brain ageing, it is not clear how the model would perform in clinical settings, as all samples used in this study are from healthy individuals without abnormalities.

Additionally, the lower brain segment was selected as ROI due to its inclusion of key structures related to ageing (e.g., hippocampus, cerebellum, amygdala) and its computational efficiency. However, no systematic evaluation of the individual contributions of these regions was performed. Future investigations could focus on the predictive significance of smaller subregions within the ROI to further refine the cropping strategy. Although our study highlights the capability of the CNN model to achieve high accuracy with just 1 million parameters, the exploration of alternative methods, such as State-Space Models (SSMs), may offer significant advantages. SSMs are particularly effective in capturing long-range dependencies and global context across brain regions, enabling them to model interactions between distant areas of the brain that contribute to ageing. This could provide deeper insight into the brain regions that most strongly influence the ageing process. A promising future direction would be the integration of CNNs with SSMs to develop lightweight models that combine CNNs' ability to extract localised spatial features with the ability of SSMs to capture global patterns. Such hybrid models could offer improved accuracy and interpretability, which are crucial for clinical applications.

Further, although we demonstrated similar performance on external datasets compared to internal ones, this study has additional limitations. Our method relies on adapting the external domain to match the source domain and benefits from fine-tuning the model based on external data. There is still a need to explore techniques that allow models to be directly applied to external clinical applications without requiring domain adaptation or fine-tuning.

In conclusion, we presented a low-complexity, accurate CNN model capable of estimating precise brain age based on MR scans acquired from different sites with minimal preprocessing. This study significantly improved training speed by using only a segment of the brain while maintaining accuracy comparable to whole-brain scans, demonstrating the adaptability of small-scale models to data from external, unseen sites.

## VII. DATA AVAILABILITY STATEMENT

The participants used for this study come from publicly available datasets and can be accessed through: Autism Brain Imaging Data Exchange I (ABIDE I, http://fcon_1000.projects.nitrc.org/indi/abide/abide_I/), Autism Brain Imaging Data Exchange II (ABIDE II, http://fcon_1000.projects.nitrc.org/indi/abide/abide_II/), Beijing Normal University (BNU,http://fcon_1000.projects. nitrc.org/indi/retro/BeijingEnhanced/), Berlin (Berlin, https//fcon_1000.projects.nitrc.org/) Cleveland Clinic (Cleveland CCF, https://fcon_1000.projects.nitrc.org/), Dallas Lifespan Brain Study (DLBS, https://fcon_1000.projects.nitrc.org/indi/retro/), Alzheimer'sDisease Neuroimaging Initiative database(ADNI, https://adni.loni.usc.edu),Information eXtraction from Images (IXI, https://brain- development.org/), and Train (Train-39, http://fcon_1000.projects. nitrc.org/indi/retro/Train-39/).

## REFERENCES


[1] T. Vos, A. D. Flaxman, M. Naghavi, R. Lozano, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, S. Abdalla, and V. Aboyans, "Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the global burden of disease study 2010," *lancet*, vol. 380, no. 9859, pp. 2163–2196, Dec. 2012.

[2] S. Chatterjee, K. Steinhaeuser, A. Banerjee, S. Chatterjee, and A. Ganguly, "Sparse group lasso: Consistency and climate applications," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2012, pp. 47–58.

[3] L. J. Seidman, E. M. Valera, and G. Bush, "Brain function and structure in adults with attention-deficit/hyperactivity disorder," *Psychiatric Clinics*, vol. 27, no. 2, pp. 323–347, Jun. 2004.

[4] C.-C. Ng, M. H. Yap, Y.-T. Cheng, and G.-S. Hsu, "Hybrid ageing patterns for face age estimation," *Image Vis. Comput.*, vol. 69, pp. 92–102, Jan. 2018.

[5] H. Y. Lee, S.-E. Jung, Y. N. Oh, A. Choi, W. I. Yang, and K.-J. Shin, "Epigenetic age signatures in the forensically relevant body fluid of semen: A preliminary study," *Forensic Sci. Int., Genet.*, vol. 19, pp. 28–34, Nov. 2015.

[6] K. Armanious, S. Abdulatif, W. Shi, S. Salian, T. Küstner, D. Weiskopf, T. Hepp, S. Gatidis, and B. Yang, "Age-Net: An MRI-based iterative framework for brain biological age estimation," *IEEE Trans. Med. Imag.*, vol. 40, no. 7, pp. 1778–1791, Jul. 2021.

[7] C. Gaser, K. Franke, S. Klöppel, N. Koutsouleris, and H. Sauer, "BrainAGE in mild cognitive impaired patients: Predicting the conversion to Alzheimer's disease," *PLoS ONE*, vol. 8, no. 6, Jun. 2013, Art. no. e67346.

[8] S. A. Valizadeh, J. Hänggi, S. Mérillat, and L. Jäncke, "Age prediction on the basis of brain anatomical measures," *Human Brain Mapping*, vol. 38, no. 2, pp. 997–1008, Feb. 2017.

[9] J. H. Cole, S. J. Ritchie, M. E. Bastin, M. C. V. Hernández, S. M. Maniega, N. A. 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. Sharp, and I. J. Deary, "Brain age predicts mortality," *Mol. Psychiatry*, vol. 23, no. 5, pp. 1385–1392, Apr. 2017.

[10] H. Sajedi and N. Pardakhti, "Age prediction based on brain MRI image: A survey," *J. Med. Syst.*, vol. 43, no. 8, p. 279, Aug. 2019.

[11] T.-W. Huang, H.-T. Chen, R. Fujimoto, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "Age estimation from brain MRI images using deep learning," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 849–852.

[12] K. Ito, R. Fujimoto, T.-W. Huang, H.-T. Chen, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "Performance evaluation of age estimation from T1-weighted images using brain local features and CNN," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 694–697.

[13] J. H. Cole, R. P. K. Poudel, D. Tsagkrasoulis, M. W. A. 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, pp. 115–124, Dec. 2017.

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

[15] K. R. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, May 2016.

[16] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, Sep. 2011.

[17] C. D. Good, I. S. Johnsrude, J. Ashburner, R. N. A. Henson, K. J. Friston, and R. S. J. Frackowiak, "A voxel-based morphometric study of ageing in 465 normal adult human brains," *NeuroImage*, vol. 14, no. 1, pp. 21–36, Jul. 2001.

[18] L. Su, L. Wang, H. Shen, and D. Hu, "Age-related classification and prediction based on MRI: A sparse representation method," *Proc. Environ. Sci.*, vol. 8, pp. 645–652, Jan. 2011.

[19] Y. Ge, R. I. Grossman, J. S. Babb, M. Rabin, L. J. Mannon, and D. L. Kolson, "Age-related total gray matter and white matter changes in normal adult brain. Part I: Volumetric MR imaging analysis," *Amer. J. Neuroradiol.*, vol. 23, no. 8, pp. 1327–1333, Sep. 2002.

[20] W. Gong, C. F. Beckmann, A. Vedaldi, S. M. Smith, and H. Peng, "Optimising a simple fully convolutional network for accurate brain age prediction in the PAC 2019 challenge," *Frontiers Psychiatry*, vol. 12, May 2021, Art. no. 627996.

[21] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, Dec. 2018, pp. 2488–2498.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2015, pp. 448–456.

[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[24] M. I. Sereno, J. Diedrichsen, M. Tachrount, G. Testa-Silva, H. d'Arceuil, and C. De Zeeuw, "The human cerebellum has almost 80% of the surface area of the neocortex," *Proc. Nat. Acad. Sci. USA*, vol. 117, no. 32, pp. 19538–19543, Aug. 2020.

[25] N. Raz and K. M. Rodrigue, "Differential aging of the brain: Patterns, cognitive correlates and modifiers," *Neurosci. Biobehavioral Rev.*, vol. 30, no. 6, pp. 730–748, Jan. 2006.

[26] N. Raz, "The aging brain observed in vivo: Differential changes and their modifiers," in *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging*, R. Cabeza, Ed., New York, NY, USA: Oxford Univ. Press, 2004, pp. 19–57.

[27] A. M. Fjell, K. B. Walhovd, C. Fennema-Notestine, L. K. McEvoy, D. J. Hagler, D. Holland, J. B. Brewer, and A. M. Dale, "One-year brain atrophy evident in healthy aging," *J. Neurosci.*, vol. 29, no. 48, pp. 15223–15231, Dec. 2009.

[28] N. Raz, U. Lindenberger, K. M. Rodrigue, K. M. Kennedy, D. Head, A. Williamson, C. Dahle, D. Gerstorf, and J. D. Acker, "Regional brain changes in aging healthy adults: General trends, individual differences and modifiers," *Cerebral Cortex*, vol. 15, no. 11, pp. 1676–1689, Nov. 2005.

[29] A.-T. Du, N. Schuff, L. L. Chao, J. Kornak, W. J. Jagust, J. H. Kramer, B. R. Reed, B. L. Miller, D. Norman, H. C. Chui, and M. W. Weiner, "Age effects on atrophy rates of entorhinal cortex and hippocampus," *Neurobiol. Aging*, vol. 27, no. 5, pp. 733–740, May 2006.

[30] H. Jiang, N. Lu, K. Chen, L. Yao, K. Li, J. Zhang, and X. Guo, "Predicting brain age of healthy adults based on structural MRI parcellation using convolutional neural networks," *Frontiers Neurol.*, vol. 10, p. 1346, Jan. 2020.

[31] H. Jiang, J. Guo, H. Du, J. Xu, and B. Qiu, "Transfer learning on T1-weighted images for brain age estimation," *Math. Biosciences Eng.*, vol. 16, no. 5, pp. 4382–4398, 2019.

[32] X. Li, Z. Hao, D. Li, Q. Jin, Z. Tang, X. Yao, and T. Wu, "Brain age prediction via cross-stratified ensemble learning," *NeuroImage*, vol. 299, Oct. 2024, Art. no. 120825.

[33] A. G. de Lange, M. Anatürk, J. Rokicki, L. K. M. Han, K. Franke, D. Alnæs, K. P. Ebmeier, B. Draganski, T. Kaufmann, L. T. Westlye, T. Hahn, and J. H. Cole, "Mind the gap: Performance metric evaluation in brain-age prediction," *Hum. Brain Mapping*, vol. 43, no. 10, pp. 3113–3129, Jul. 2022.

[34] V. M. Bashyam, "MRI signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide," *Brain*, vol. 143, no. 7, pp. 2312–2324, Jul. 2020.

[35] G. Levakov, G. Rosenthal, I. Shelef, T. R. Raviv, and G. Avidan, "From a deep learning model back to the brain—Identifying regional predictors and their relation to aging," *Hum. Brain Mapping*, vol. 41, no. 12, pp. 3235–3252, Aug. 2020.

[36] S. He, D. Pereira, J. David Perez, R. L. Gollub, S. N. Murphy, S. Prabhu, R. Pienaar, R. L. Robertson, P. Ellen Grant, and Y. Ou, "Multi-channel attention-fusion neural network for brain age estimation: Accuracy, generality, and interpretation with 16,705 healthy MRIs across lifespan," *Med. Image Anal.*, vol. 72, Aug. 2021, Art. no. 102091.

[37] S. He, P. E. Grant, and Y. Ou, "Global-local transformer for brain age estimation," *IEEE Trans. Med. Imag.*, vol. 41, no. 1, pp. 213–224, Jan. 2022.

[38] N. K. Dinsdale, E. Bluemke, S. M. Smith, Z. Arya, D. Vidaurre, M. Jenkinson, and A. I. L. Namburete, "Learning patterns of the ageing brain in MRI using deep convolutional networks," *NeuroImage*, vol. 224, Jan. 2021, Art. no. 117401.

[39] M. Ueda, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "An age estimation method using 3D-CNN from brain MRI images," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 380–383.

[40] B. A. Jonsson, G. Bjornsdottir, T. E. Thorgeirsson, L. M. Ellingsen, G. B. Walters, D. F. Gudbjartsson, H. Stefansson, K. Stefansson, and M. O. Ulfarsson, "Brain age prediction using deep learning uncovers associated sequence variants," *Nature Commun.*, vol. 10, no. 1, p. 5409, Nov. 2019.

[41] X. Feng, Z. C. Lipton, J. Yang, S. A. Small, and F. A. Provenzano, "Estimating brain age based on a uniform healthy population with deep learning and structural magnetic resonance imaging," *Neurobiol. Aging*, vol. 91, pp. 15–25, Jul. 2020.

[42] J. Cheng, Z. Liu, H. Guan, Z. Wu, H. Zhu, J. Jiang, W. Wen, D. Tao, and T. Liu, "Brain age estimation from MRI using cascade networks with ranking loss," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3400–3412, Dec. 2021.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

**FATMA ELTASHANI** received the B.Eng. degree in electrical and electronic engineering from The University of Manchester, in 2018. She is currently a Ph.D. Researcher with the Department of Electrical and Electronic Engineering, The University of Manchester, exploring the use of AI systems for the prediction of cognitive aging to help early detection of neurodegenerative diseases. She has been researching to find possible ways to reduce the computational complexity of deep learning age prediction systems. Her research interests include cognitive aging, deep learning, machine learning, and MRI imaging.

**MARIO PARRENO-CENTENO** is currently a Research Associate with the Cancer Bioinformatics Group, King's College London, and an Honorary Research Fellow with the UCL Genetics Institute, Department of Genetics, Evolution, and Environment. With a strong background in bioinformatics and machine learning, his research focuses on using computational techniques to model multi-omics data to understand the systemic changes in cells that may lead to cancer initiation, progression, and resistance to therapies. He is a Data Scientist with a Ph.D. from the Centre for Doctoral Training in Cloud Computing for Big Data, Newcastle University. During his Ph.D. studies, he focused on applying machine learning techniques to model data from diverse sources, including time series and imaging data.

**JAMES H. COLE** is currently a Professor of neuroimage computing with the Centre for Medical Imaging Computer (CMIC) and Dementia Research Centre (DRC), University College London (UCL). His work uses machine learning, deep learning, and related statistical methods with the goal of developing clinically useful neuroimaging tools. He is a Principal Investigator with the MANIFOLD Laboratory. His research interests include brain aging, neurological and psychiatric diseases, with a particular focus on aging, neurodegeneration, and dementia.

**JOÃO PAULO PAPA** (Senior Member, IEEE) is currently an Associate Professor with the Department of Computing, São Paulo State University, Brazil. He was a Visiting Scholar with Harvard University, USA, from 2014 to 2015. He is an Active Member of both Brazilian and Latin American communities for *Pattern Recognition* and related fields. He is a Fellow Member of the Alexander von Humboldt Foundation, Germany, and the Brazilian National Council for Scientific and Technological Development. He is currently the Brazilian Chair of the International Association of Pattern Recognition and the Chair of the IEEE Task Force on Business Intelligence and Knowledge Management.

**FUMIE COSTEN** (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering and the Ph.D. degree in informatics from Kyoto University, Japan. From 1993 to 1997, she was with the Advanced Telecommunication Research International, Kyoto, where she was engaged in research on direction-of-arrival estimation based on multiple signal classification algorithm for 3-D laser microvision. She filed three patents from the research in Japan, in 1999. She was invited to give five talks in Sweden and Japan, from 1996 to 2014. From 1998 to 2000, she was with Manchester Computing, The University of Manchester, U.K., where she was engaged in research on metacomputing and has been a Lecturer, since 2000. Her research interests include computational electromagnetics in such topics as a variety of finite difference time domain methods for microwave frequency range, microwave imaging, application of deep learning to FDTD, radar images and medical images, clinical information, CSF/blood biomarkers, and RNA. She filed a patent for the research on boundary conditions in USA, in 2012. Her work extends to the hardware acceleration of the computation using general-purpose computing on graphics processing units, streaming single instruction multiple data extension, and advanced vector eXtentions instructions. She received an ATR Excellence in Research Award, in 1996, the Best Paper Award from the 8th International Conference on High Performance Computing and Networking Europe, in 2000, and the Teaching Excellence Award 2021.