



Brain Age Prediction of Children Using Routine Brain MR Images via Deep Learning

Jin Hong^{1,2†}, Zhangzhi Feng^{2†}, Shui-Hua Wang^{3,4}, Andrew Peet⁵, Yu-Dong Zhang^{1,6*}, Yu Sun^{5,7*} and Ming Yang^{2*}

¹ School of Informatics, University of Leicester, Leicester, United Kingdom, ² Department of Radiology, Children's Hospital of Nanjing Medical University, Nanjing, China, ³ School of Architecture Building and Civil Engineering, Loughborough University, Loughborough, United Kingdom, ⁴ School of Mathematics and Actuarial Science, University of Leicester, Leicester, United Kingdom, ⁵ Institute of Cancer & Genomic Science, University of Birmingham, Birmingham, United Kingdom, ⁶ Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, ⁷ International Laboratory for Children's Medical Imaging Research, School of Biology Science and Medical Engineering, Southeast University, Nanjing, China

OPEN ACCESS

Edited by:

Kirsten A. Donald,
University of Cape Town, South Africa

Reviewed by:

Jonathan Ipser,
University of Cape Town, South Africa
Maurizio Elia,
Oasi Research Institute (IRCCS), Italy

*Correspondence:

Yu-Dong Zhang
yudongzhang@ieee.org
Yu Sun
sunyu@seu.edu.cn
Ming Yang
yangming19710217@163.com

†These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Pediatric Neurology,
a section of the journal
Frontiers in Neurology

Received: 17 July 2020

Accepted: 04 September 2020

Published: 19 October 2020

Citation:

Hong J, Feng Z, Wang S-H, Peet A,
Zhang Y-D, Sun Y and Yang M (2020)
Brain Age Prediction of Children Using
Routine Brain MR Images via Deep
Learning. *Front. Neurol.* 11:584682.
doi: 10.3389/fneur.2020.584682

Predicting brain age of children accurately and quantitatively can give help in brain development analysis and brain disease diagnosis. Traditional methods to estimate brain age based on 3D magnetic resonance (MR), T1 weighted imaging (T1WI), and diffusion tensor imaging (DTI) need complex preprocessing and extra scanning time, decreasing clinical practice, especially in children. This research aims at proposing an end-to-end AI system based on deep learning to predict the brain age based on routine brain MR T1-weighted images of healthy children aged 0 to 5 years old and randomly divided those images into training data including 176 subjects and test data including 44 subjects. Data augmentation technology, which includes scaling, image rotation, translation, and gamma correction, was employed to extend the training data. A 10-layer 3D convolutional neural network (CNN) was designed for predicting the brain age of children and it achieved reliable and accurate results on test data with a mean absolute deviation (MAE) of 67.6 days, a root mean squared error (RMSE) of 96.1 days, a mean relative error (MRE) of 8.2%, a correlation coefficient (R) of 0.985, and a coefficient of determination (R^2) of 0.971. Specially, the performance on predicting the age of children under 2 years old with a MAE of 28.9 days, a RMSE of 37.0 days, a MRE of 7.8%, a R of 0.983, and a R^2 of 0.967 is much better than that over 2 with a MAE of 110.0 days, a RMSE of 133.5 days, a MRE of 8.2%, a R of 0.883, and a R^2 of 0.780.

Keywords: magnetic resonance imaging, deep learning, brain age, convolutional neural network, artificial intelligence

INTRODUCTION

The brain development of children undergoes a rapid and complex process, especially in the first 2 years after birth (1, 2). The early brain development follows the law of myelination from caudal to rostral, posterior to anterior regions, central to peripheral locations, which is closely related to the development of sensory, motor, and cognitive ability (3). Delayed brain development can lead to

intellectual disability, language disorder, activity limitation, and other manifestations in children, which seriously affect their quality of life. Therefore, accurate and quantitative evaluation of brain development, early identification, and intervention treatment is particularly important for children with brain development analysis and brain disease diagnosis.

At present, brain magnetic resonance (MR) imaging is a reliable method to evaluate brain development (brain age) due to its non-invasive, high soft tissue resolution and multi-parameter imaging advantages. Recently, the main ways of MR image to evaluate brain development are as follows: morphometry [including measurement of brain volume (4–6), cortical thickness (7), surface area (7, 8), etc.], white matter diffusion (9, 10), functional connectivity (11–14). However, there are some drawbacks within these studies: the need of some special sequences with long scanning time, complex data post-processing, and group-level comparison results without quantitative analysis to individuals, which limit their wide use in clinical situations.

With the development of deep learning, more and more sophisticated deep neural networks have been proposed to analysis massive image, voice, or video data. Of these, convolutional neural network (CNN) of deep learning has achieved great success with superior performance beyond human experts in many computer vision and speech recognition tasks since it was put forward (15–20). In the field of medical image analysis, CNN-based method has been also proposed for disease diagnosis and lesion detection with high performance in accuracy, such as the classification and detection of lung nodules (21, 22), the recognition of melanoma (23), the detection of cerebral microbleeds (24–26), as well as the classification of Alzheimer's disease (27, 28). In addition, brain age prediction based CNN model has been proved to be a reliable and heritable biomarker of brain aging and can be used to indicate the risk of brain degenerative diseases (29, 30), whereas it has not been reported in young children up to now. Furthermore, unlike traditional machine learning approaches that implement feature extraction, feature reduction, and classification separately, CNN combines them as an end-to-end system, from raw images to the corresponding target values, avoiding complicated image preprocessing and manual design of appropriate features. The excellent performance and transferability lead us to believe that CNN-based method should be the most promising resolution for most clinical applications, including brain age prediction of children.

In this paper, we collected 220 routine brain MR images of healthy children for investigating the brain age of children based on deep learning. Data augmentation was utilized to extend the training data for avoiding the potential over-fitting and enhancing generalizability of the model. With delicate design of structure and careful setting of hyper parameters, we proposed a 3D deep neural network and achieved high performance. We analyzed the prediction results of different age groups in detail and compared them with those of other two state-of-the-art methods. The factors in the proposed model that may affect the prediction results were investigated comprehensively. Furthermore, we compared the proposed 3D CNN with the

corresponding 2D CNN that has a similar structure in predicting brain age of children with 3D MR image data.

METHODS

Dataset Acquisition

Ethical approval for the research was obtained from the ethics committee of Children's Hospital of Nanjing Medical University. This is a retrospective study, and informed written consent was thus waived. The dataset consists of T1-weighted images of 220 healthy children aged 0 to 5 years old. The data were all acquired using a 1.5T Siemens Avanto Scanner, but scanning parameters of newborns (≤ 1 month) are different from older children due to variation in water content of brain tissue. Scans of newborns were imaged using a T1-weighted spin-echo sequence (repetition time [TR] = 4,490 ms, echo time [TE] = 7.5 ms, flip angle [FA] = 150° , 18 slices, slice thickness = 4.5 mm, FOV = 180×180 mm, voxel dimensions = $1.0 \times 0.7 \times 4.5$ mm). Scans of older children (> 1 month) were also imaged using the T1-weighted spin-echo sequence (TR = 3,850 ms, TE = 7.3 ms, FA = 150° , 22 slices, slice thickness = 5.0 mm, FOV = 220×220 mm, voxel dimensions = $1.4 \times 1.0 \times 5.0$ mm). Those whose brain MR image quality was good enough to diagnose and reports were diagnosed as normal by two experienced radiologists, and whose history, clinical data, and phone call following-up can't show the existence of neurological disease were enrolled into our dataset. Premature infants, subjects who were diagnosed with congenital diseases (congenital heart diseases, Down's syndrome, etc.), neurodevelopmental or mental disorders (neurodevelopmental delay, autism, etc.), and other serious illnesses (hypoxic-ischemic encephalopathy, cerebral hemorrhage, septicopyemia, etc.) affecting brains were excluded from our dataset. Furthermore, we used downsampling method to convert the stacked 2D brain MR images of newborns and older children to the same size of $128 \times 116 \times 12$.

Data Augmentation Technology

Basically, data augmentation methods are extensively used to train a deep neural network having huge parameters for improving prediction accuracy that had been validated in the "Results" section. In our experiments, four commonly used methods of data augmentation were employed to enhance the training dataset. They are listed as: (a) scaling, (b) image rotation, (c) translation, and (d) gamma correction. We scaled images with scaling factor of 0.85 to 1.15 with step of 0.03 for generating 10 new images. Image rotation was used to generate 10 new images with rotation angle of -15 to 15 degrees increased by 3 degrees. We translated images with factor of -0.1 to 0.1 with step of 0.02 diagonally for generating 10 new images. Gamma correction with gamma value of 0.85 to 1.15 increased by 0.03 was employed to generate 10 new images. At last, we augmented the training dataset by 41 times using data augmentation methods.

Noted: For one routine brain MR image with size of $128 \times 116 \times 12$ in this paper, we split the volume into 12 slices, then used the same transformation method to process every slice, and finally stacked those slices into a 3D image (see **Figure 1**). We achieved the transformation of the whole 3D image by this way.

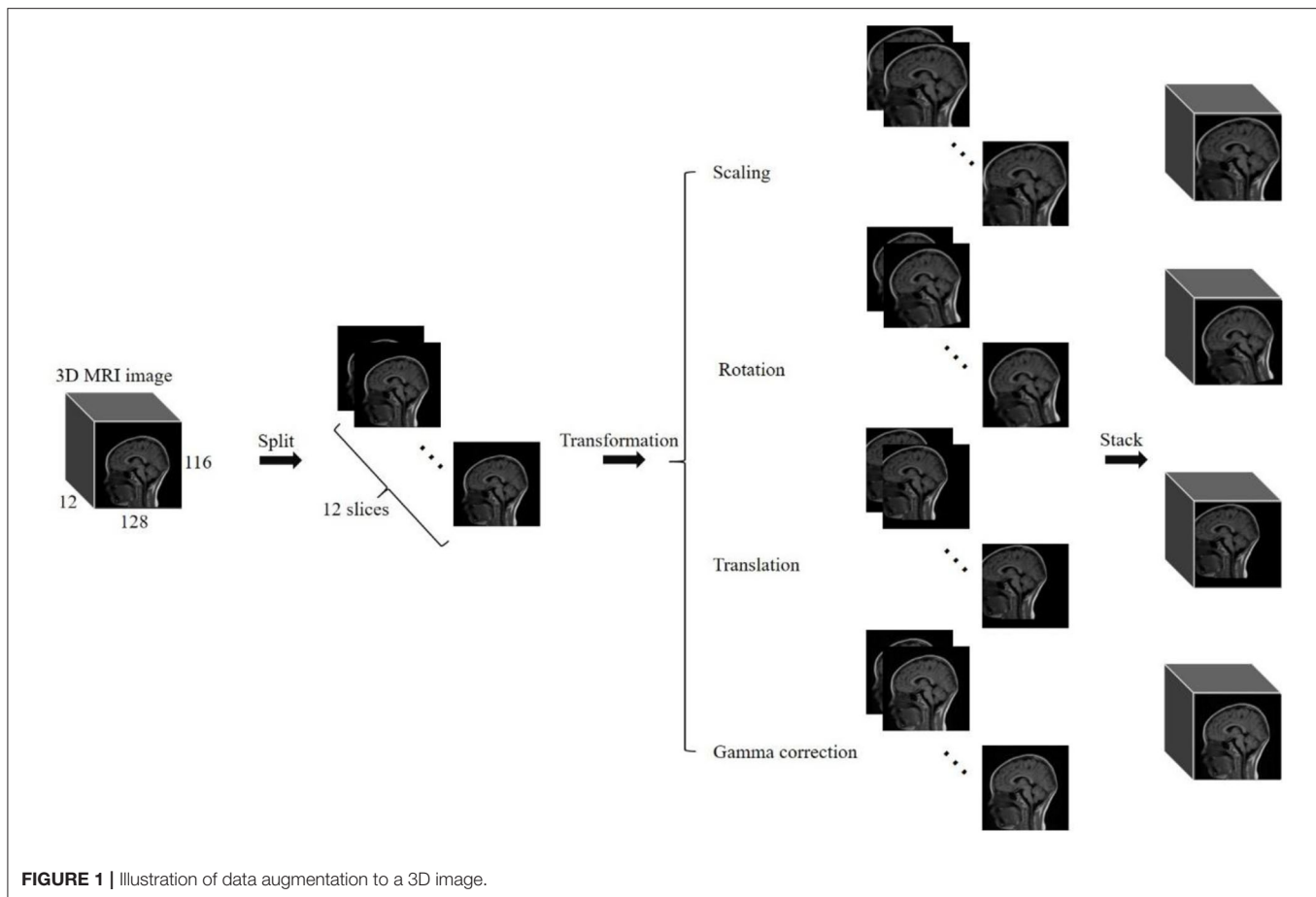


FIGURE 1 | Illustration of data augmentation to a 3D image.

Proposed 3D CNN Architecture

A 3D CNN was proposed to predict the brain age of children using brain MR images with size of $128 \times 116 \times 12$. The 3D image was input to the model and then a single scalar denoting the brain age was output. The proposed 3D CNN model, shown in **Figure 2**, contains 7 3D convolution layers, 4 3D max pooling layers, and 3 fully connected layers. All convolution layers are followed by 3D batch-normalization (31) and ReLU activation function (32), while the first two fully connected layers are followed by ReLU activation function. All convolution layers have the same kernel size of $3 \times 3 \times 3$, stride size of $1 \times 1 \times 1$, and padding size of $1 \times 1 \times 1$, which means the feature map size is the same as that of the input. The kernel size in the first, second, third, and fourth max pooling layer are $4 \times 4 \times 1$, $3 \times 3 \times 1$, $3 \times 3 \times 3$, and $3 \times 3 \times 3$, respectively, and the stride size is equal to the kernel size in all max pooling layers.

The mean absolute error (MSE) was used as the loss function. The reliable and commonly used stochastic gradient descent with momentum of 0.9 (SGDM) was employed as the optimization method. The mini-batch and epoch were set to 64 and 40, respectively. The initial learning rate was set to 0.0000008 and decreased by 10% every epoch. The weights were initialized randomly.

Note: 10 runs were implemented for accounting for the stochastic properties of CNN, and the

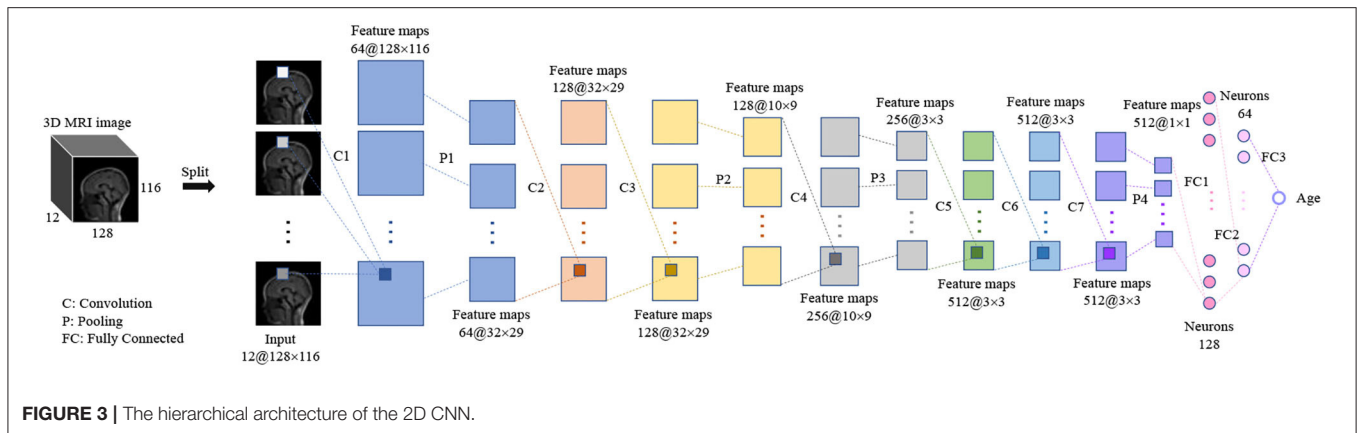
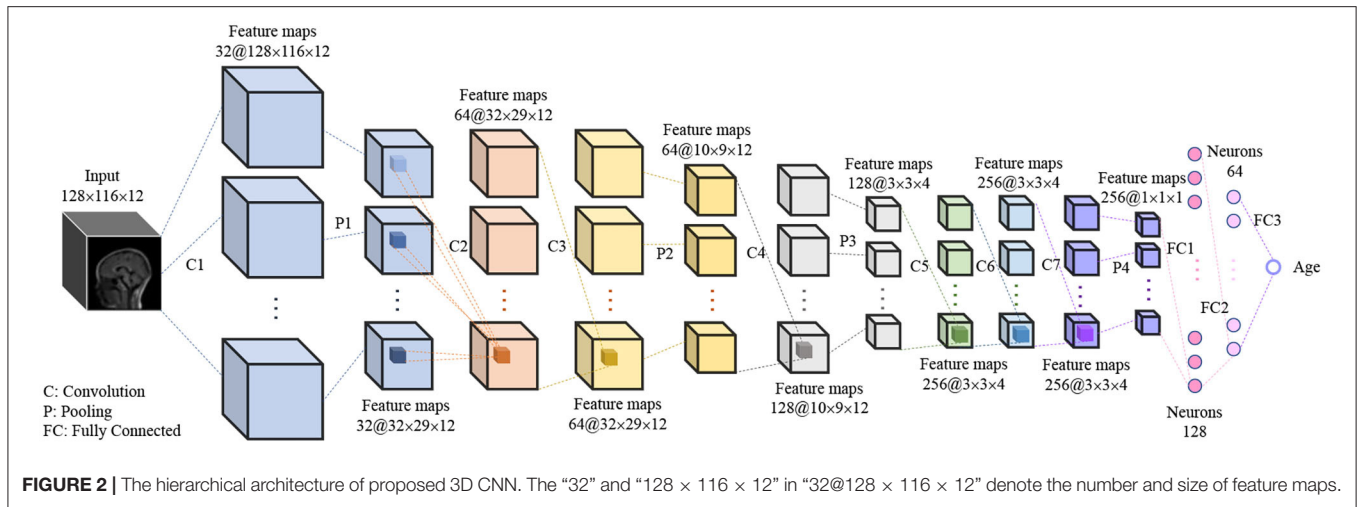
average value of the 10 runs is regarded as the final result.

2D CNN Architecture

We designed a 2D CNN model, shown in **Figure 3**, according to the structure of the proposed 3D CNN, so both have similar hierarchical structures. The 3D image with size of $128 \times 116 \times 12$ was split into 12 slices, and those slices were then input to the 2D CNN model, and finally the age was given. Same as the proposed 3D CNN, all convolution layers in 2D CNN are followed by batch normalization and ReLU, and the first two fully connected layers are followed by ReLU. The kernel size, stride size, and padding size in convolution layers are 3×3 , 1×1 , and 1×1 , respectively. The kernel size is equal to stride size in all max pooling layers, and they are 4×4 , 3×3 , 3×3 , and 3×3 , respectively, from the first to the last max pooling layer. In terms of hyperparameters, the loss function, optimizer, mini-batch, epoch, learning rate, and weights initialization were set the same as the proposed 3D CNN.

Software Availability and PC Configuration

All data augmentation methods were implemented using `imgaug` (<https://github.com/aleju/imgaug>). All experiments of deep learning were carried out on PyTorch (<https://pytorch.org/>). The running environment of the programs: i9-9900k CPU, NVIDIA GeForce RTX 2080 Ti GPU, and 16.0 GB RAM.



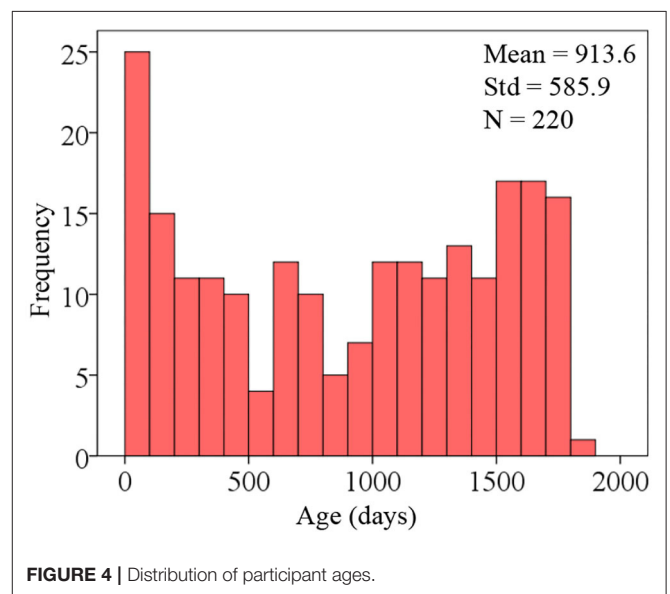
RESULTS

Dataset Characteristics

To develop an AI system for predicting the brain age of children using routine clinical brain MR images, we enrolled 220 subjects aged 0 to 5 years old (Figure 4 shows the distribution of participant ages with 100-day intervals) and scanned them to achieve the brain MR images. The hold-out method was employed to divide the 220-image dataset into two parts randomly, and one part containing 176 images (80%) was regarded as training dataset and the other part containing 44 images (20%) as test dataset. The reason for abandoning the validation dataset is that the whole dataset only contains 220 subjects. Table 1 gives the demographic information of the training and test datasets. Since the amount of the training dataset is slightly small to train a deep neural network, data augmentation was implemented for generating new “fake” images. At last, the training dataset containing 7,216 images and the test dataset containing 44 images were obtained.

Performance of the Proposed Model

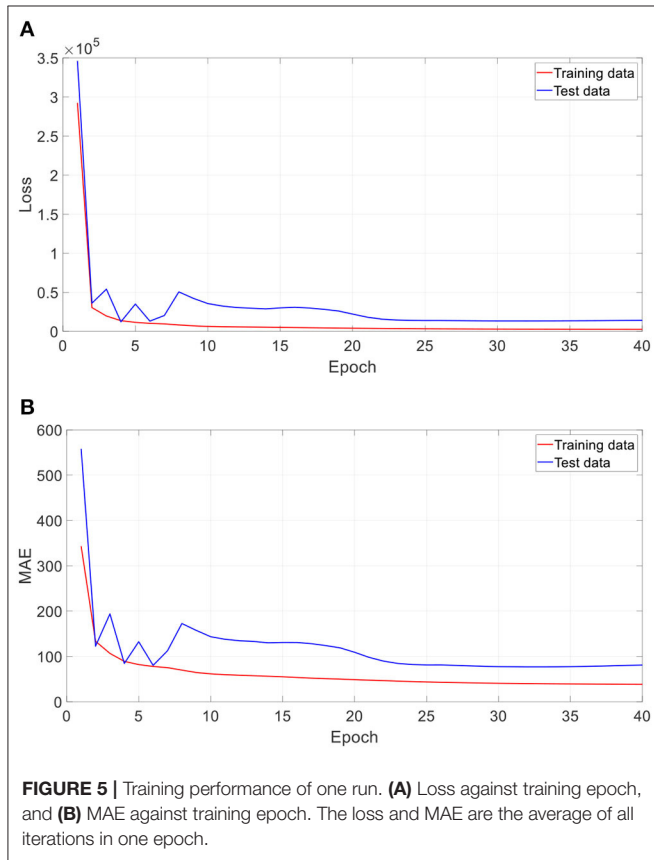
With grid search and trial-and-error methods, we optimized a 3D CNN model for predicting the children’s age more accurately and reliably. The detailed information of the proposed model can be



found in Figure 2. The model was trained by the data-augmented dataset including 7,216 images and evaluated on the test dataset including 44 images. In our experiments, the learnable weights

TABLE 1 | Subjects demographic (Std denotes standard deviation).

	Total dataset		Test dataset		Training dataset	
	0–2 years old	2–5 years old	0–2 years old	2–5 years old	0–2 years old	2–5 years old
Subjects	88	132	23	21	65	111
Age (days)	4–697	731–1820	36–680	749–1687	4–697	731–1820
Mean ± Std	283.6 ± 215.7	1333.5 ± 314.2	244.2 ± 201.0	1267.0 ± 288.4	297.6 ± 220.5	1346.1 ± 318.4



of model were initialized randomly, and the random seed was not fixed, causing the randomness of prediction results. Thus, we implemented 10 runs under the same settings of the model for ensuring the reliability of the results.

Figure 5 shows the training performance of one typical run. As **Figure 5** shows, after 40 epochs (iterations through the whole training dataset), both training dataset and test dataset in loss and MAE reached a plateau and were at a minimum, which means that the training process has converged.

Figure 6 shows the average and standard deviation of prediction results of 10 runs under the same setting. It is found that most true data fall within the standard deviation of predicted data, which means that the predictions can fit the true data well. To further quantitatively evaluate the prediction accuracy of the model, MAE, RMSE, MRE, *R*, and *R*² between the average values and the true values were employed (**Table 2**). With a MAE of 67.6

days, a RMSE of 96.1 days, a MRE of 8.2%, a *R* of 0.985, and a *R*² of 0.971, the proposed model was considered to achieve quite high accuracy in predicting the brain age of children aged 0 to 5 years old.

Since the brain development of infants under 2 years old is heterogeneous and particularly rapid, it is necessary to divide the age into two age groups according to 2 years old and evaluate the prediction results of the two groups separately. **Table 2** gives the assessment results. We found that age predictions for children under 2 years old are significantly better than those over 2 according to all evaluation indicators. We can also observe that most predictions under 2 years old are closer to the true values compared with those over 2 years old in **Figure 6**. Comparing to the 0–2 age group, there is a stronger correlation between predicted and true values in the age group from 0 to 5 years old according to *R*, but there is a bigger MAE. The reason is that the true values in 0–2 age group are smaller as a whole than that in 0–5.

To further assess the reliability of the predicted results, we gave the residual plot (Bland-Altman plot) and performed paired samples *T*-test. The residual plot was employed to show the relationship between mean and difference of the predicted and actual value, which is show in **Figure 7**. The *P*-values of the paired samples *T*-test were 0.5665, 0.9407, and 0.7979 in 0–2, 2–5, and 0–5 age groups, respectively, showing that there are no significant statistical differences between the predicted and the actual values of all age groups. As the **Figure 7A** indicating the 0–2 age group shows, 95.7% (22/23) of the points fall within the 95% limits of agreement, and the mean of difference is 4.6, which is close to 0. Similar to the 0–2 age group, 95.2% (20/21) of the points fall within the 95% limits of agreement, and the mean of difference is –2.3 according to **Figure 7B**. In terms of 0–5 age group, which is shown in **Figure 7C**, 90.9% (40/44) of the points fall within the 95% limits of agreement, and the mean of difference is 3.8 which is quite close to 0.

Impact of Data Augmentation

In this research, the amount of the obtained 220-subject dataset is big enough to draw a conclusion of statistical analysis, but it is still insufficient for training a deep neural network with huge parameters. Many researches have reported that increasing the number of samples in training data can avoid over-fitting, enhance generalizability, and improve the performance on test set (33–37). Therefore, we performed data augmentation on a 176-subject training dataset and extended the dataset to 7,216 in our proposed method.

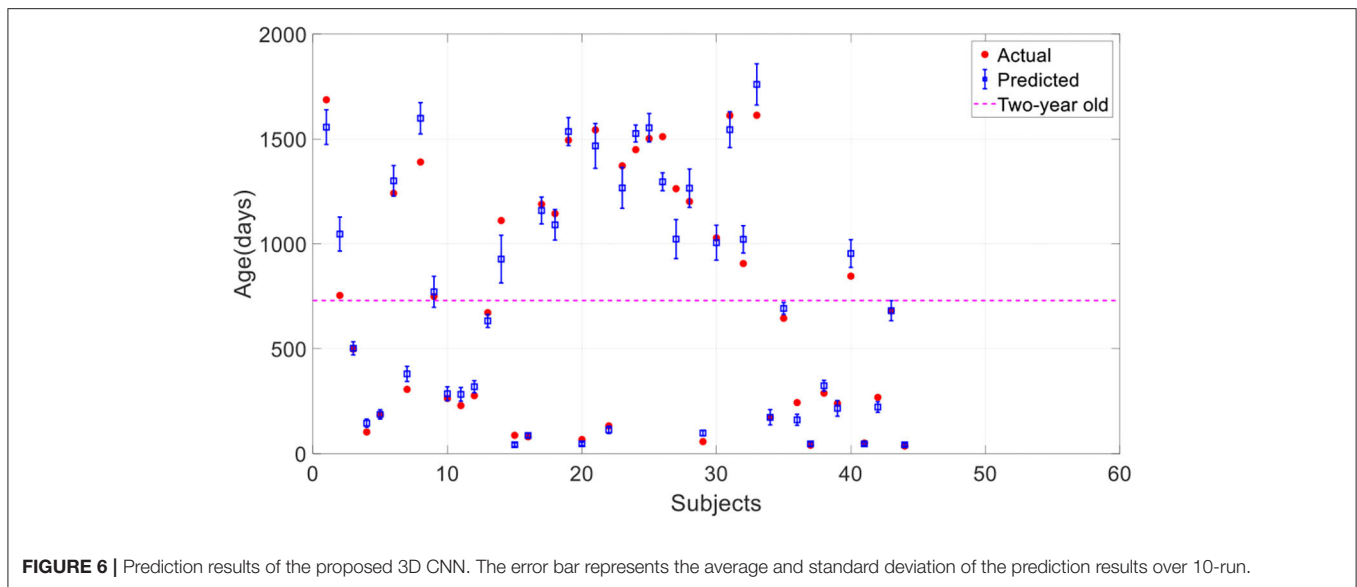


FIGURE 6 | Prediction results of the proposed 3D CNN. The error bar represents the average and standard deviation of the prediction results over 10-run.

TABLE 2 | Performance of the proposed 3D CNN in predicting children aged.

Age group (years old)	MAE (days)	RMSE (days)	MRE (%)	R	R ²
0–2	28.9	37.0	7.8	0.983	0.967
2–5	110.0	133.5	8.2	0.883	0.780
0–5	67.6	96.1	8.2	0.985	0.971

Here we investigated the impact of data augmentation on predicting children’s age using stacked 2D routine clinical brain MR images. **Figure 8** offers the prediction results of our proposed method without data augmentation. All the results are average on 10 runs. Comparing with **Figure 6**, it is found that most of the predicted values deviate from the true values further, and the standard deviation of the predicted values is larger, showing the instability of the model. **Table 3** gives a detailed comparison of our proposed method with and without data augmentation, which further confirms data augmentation can improve the prediction accuracy of the model.

Impact of Network Depth

To test the impact of network depth on performance of predicting children age, different 3D CNN including different convolution layers and fully connected layers were validated. The evaluation results are given in **Table 4**. Ten runs were implemented, and the average values were regarded as the final results. It is found that the proposed 3D CNN structure containing seven convolution layers and three fully connected layers achieved the best performance according to the comprehensive assessment of four indicators. **Figure 9** is utilized to further visualize the performance differences between different 3D CNN.

Impact of Batch Normalization

In the proposed 3D CNN structure, every 3D convolution layer is followed by a 3D batch normalization. We investigated

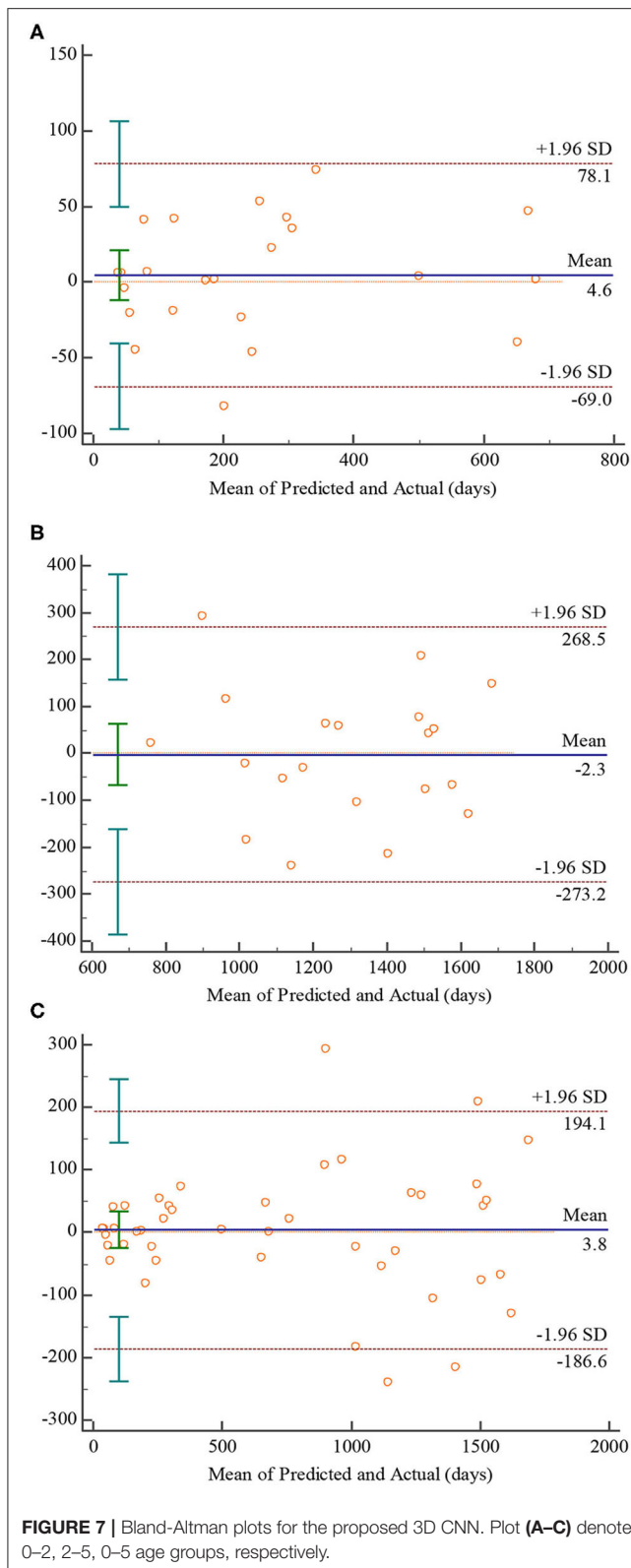
the impact of batch normalization on prediction accuracy in this section. All batch normalization layers were removed, and initial learning rate was set as 0.000000008. All other settings remain the same. **Table 5** gives the comparison result of the proposed approach with and without batch normalization. All the results are averaged on 10 runs. As **Table 5** shows, the 3D CNN without batch normalization achieved a MAE of 132.6 days, a RMSE of 189.9 days, a MRE of 15.0%, a R of 0.945, and a R² of 0.893, which is obviously worse than the proposed 3D CNN with batch normalization. **Figure 10** gives the training performance of the 3D CNN without batch normalization. As we can see, both training dataset and test dataset in loss and MAE reached the minimum plateau, indicating that the network is fully trained.

Impact of Batch Size and Learning Rate

Except for the structure, hyper parameters also can affect the 3D CNN performance. We compared different prediction results of the proposed 3D CNN trained by different batch size and initial learning rate for understanding the influence of them on the performance. **Table 6** gives the survey results. All results are average on 10 runs. As the **Table 6** shows, the 3D CNN with batch size of 64 and learning rate of 0.0000008 achieved the best prediction results according to all four evaluation indicators.

Comparing With 2D CNN

The input of 2D CNN is a 2D image with three color channels (i.e., RGB) in most natural scenes. With this regard, the simplest way for 2D CNN to deal with 3D input is to replace the color channels with the slices of the volumetric image. We designed a 2D CNN model, shown in **Figure 3**, according to the architecture of the proposed 3D CNN model for predicting the brain age of children using stacked 2D routine clinical brain MR image (gray-level) and investigated the performance differences between the two models. The comparison results are given in **Table 7**.



All the results are average on 10 runs. We observed that the proposed 3D CNN achieved better performance in terms of all the evaluation indicators.

DISCUSSION

High Reliability and Accuracy of 3D CNN for Brain Age Prediction

It is important to predict the brain age reliably and accurately for brain development analysis and brain disease diagnosis in pediatric patients. Basically, methods for predicting brain age can be divided into two categories: shallow learning algorithms and deep learning algorithms (38). So far, numerous shallow learning algorithms have been developed, such as gaussian processes regression (GPR) (29, 39, 40), support vector regression (SVR) (41, 42), partial least squares (PLS) regression (43), relevance vector regression (RVR) (44), hidden Markov model (HMM) (45), and Bayesian linear discriminant analysis (46). In terms of deep learning algorithms, CNN (29, 47) and back propagation neural network (BPNN) (48) were proposed to predict the brain age with brain MR images.

As the above references report, for achieving fairly good prediction result, all methods except CNN need to accomplish the complicated preprocessing task well including feature selection, dimension reduction, and segmentation of brain MR image into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) tissues. The manual interventions in preprocessing lead to high intra-observer and inter-observer variability, which easily biased the final interpretation. Comparing to the traditional machine learning methods, CNN-based methods are an end-to-end system that uses the raw MR image data as the input and output the age value without manual interventions, showing higher reliability and improving clinical practice (38).

Although there is no error caused by manual intervention in predicting brain age using CNN-based model, there may be systematic bias (49, 50). As reported in (49), CNN-based model will overestimate the younger and underestimate the older, decreasing the reliability of prediction results. To evaluate the reliability of the predicted results in this paper, the Bland-Altman plots characterizing the relationships between the mean and the difference of the predicted and actual value were given, showing in **Figure 7**. According to **Figures 7A,B**, the mean of difference in the 0–2 age group is slightly higher than 0, while that in the 2–5 age group is slightly lower than 0. This observation seems to indicate that the prediction results in this paper confirm the conclusion of (49). However, the means in age group of 0–2 and 2–5 are quite close to 0, and the paired samples *T*-test results revealed there are no significant statistical differences between the predicted and the actual values on both age groups, which means the predicted results of 0–2 and 2–5 age group are in good agreement with the actual age. Similarly, the predicted results of 0–5 age group are also in good agreement with the actual age according to **Figure 7C**. Therefore, the predicted results achieved by the proposed CNN-based model are considered to be reliable overall.

Furthermore, the CNN-based methods can achieve more accurate prediction results compared with traditional machine learning methods. Cole et al. (29) reported the detailed comparison between 3D CNN and GPR method in predicting the brain age using different input data (GM, WM, GM+WM, and raw data). We found that the 3D CNN achieved higher

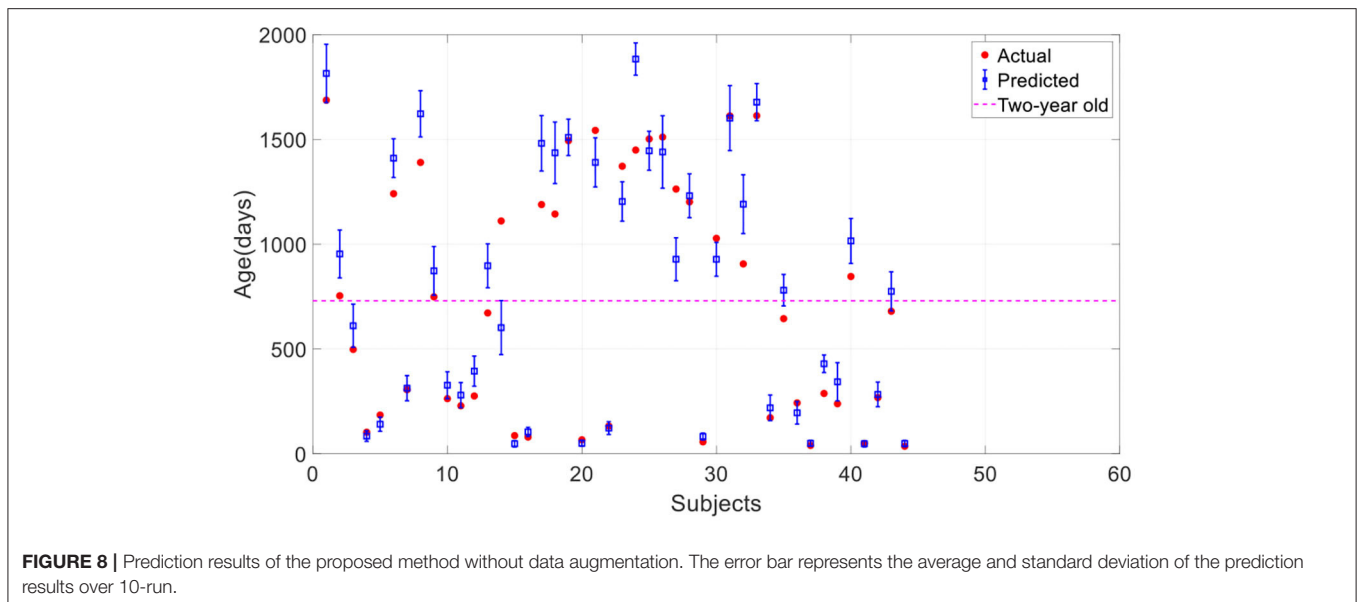


TABLE 3 | Comparison of the proposed method with and without data augmentation.

	MAE (days)	RMSE (days)	MRE (%)	R	R ²
Without data augmentation	118.3	166.7	13.7	0.963	0.926
With data augmentation	67.6	96.1	8.2	0.985	0.971

TABLE 4 | Performance of different network depths.

No. of convolution layers	No. of fully connected layers	MAE (days)	RMSE (days)	MRE (%)	R	R ²
6	2	75.9	102.8	9.2	0.983	0.967
6	3	73.8	101.6	9.0	0.984	0.968
6	4	70.1	95.1	8.4	0.985	0.971
7	2	80.4	110.5	10.0	0.981	0.962
7	3	67.6	96.1	8.2	0.985	0.971
7	4	70.3	97.1	8.6	0.985	0.970
8	2	73.4	106.8	9.2	0.982	0.964
8	3	68.7	99.5	8.2	0.984	0.969
8	4	69.5	96.6	8.3	0.985	0.971

performances than the GPR in all kinds of input data in this reference. Especially, the MAE of 4.65 years obtained by 3D CNN are much lower than the MAE of 11.81 years obtained by GPR when the raw data was used as the input.

However, it is not particularly reasonable to compare our results with the above example since the subjects used above are aged 18 to 90 years old. To the best of our knowledge, only two traditional machine learning-based methods for age prediction of young children were investigated currently. Toews et al. (51) firstly developed a feature-based developmental model

for predicting infant age using structural brain MR images. They enrolled 92 subjects aged 8–590 days and achieved a MAE of 72 days. Hu et al. (46) proposed a two-stage prediction method named Hierarchical Rough-to-Fine (HRtoF) model for predicting infant age. They enrolled 50 infants aged 14–797 days and achieved a MAE of 32.1 days. Since it is hard to collect the brain images of young children, the data amount reported in (51) and (46) is not large, <100. In our study, we spent over 5 years collecting 220 subjects, which is enough for reaching a convincing conclusion comparing to the above two studies. **Table 8** gives the performance comparison of our proposed method and the above two methods. It is found that the proposed 3D CNN gained the best performance in predicting the brain age of infant aged about 0–2 years old. The prediction accuracy of the 3D CNN for the age of 4–1,820 days is even better than the prediction accuracy of Toews’s method (51) for the age of 8–590 days.

In addition to traditional machine learning-based methods, we also compared 3D CNN with 2D CNN in predicting brain age of young children using 3D MR images. The inputted 3D images are stacked 2D brain MR images (slices) and there is a gap between two adjacent slices in the actual location of the brain. Thus, we speculated that the correlation between slices will not be great, and we think that 2D CNN model may also be able to complete the age prediction task well with the inputted 3D images. If the prediction effect of the 2D CNN is the same as that of the 3D CNN, then the 2D CNN will be more recommended in clinical practice, because the 2D model requires much less computation and computer memory. However, as **Table 7** shows, the proposed 3D CNN outperformed the 2D CNN significantly. This result shows that the small correlation between adjacent slices is beneficial to the prediction accuracy of the model, and also shows the 3D CNN employing 3D kernels is a more reliable resolution that can take all full advantage of spatial contextual information of the 3D MR images for more accurate age prediction (16, 29, 52).

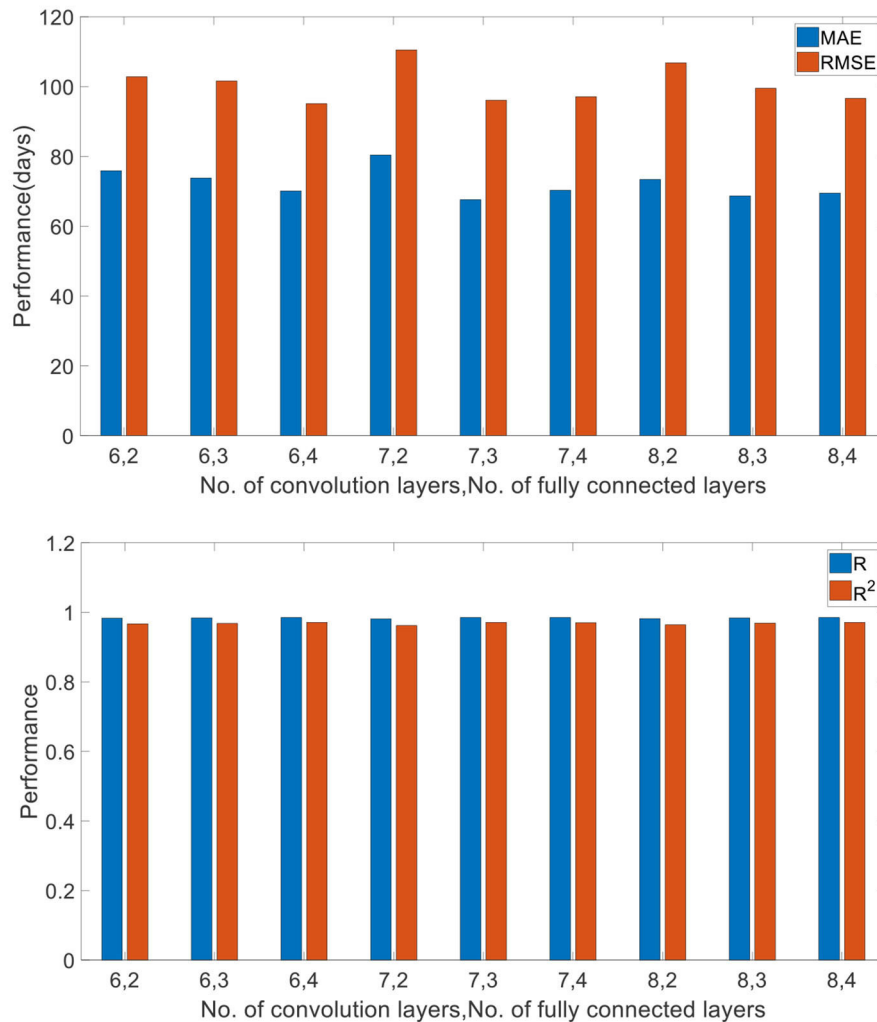


FIGURE 9 | Comparison of different network depths. “6, 2” (“No. of convolution layers, No. of fully connected layers”) denotes 6 convolution layers and 2 fully connected layers.

TABLE 5 | Comparison of the proposed method with and without batch normalization.

	MAE (days)	RMSE (days)	MRE (%)	R	R ²
Without batch normalization	132.6	189.9	15.0	0.945	0.893
With batch normalization	67.6	96.1	8.2	0.985	0.971

Predictions for Children Under 2 Years Old Are Better Than Those Over 2

As Table 2 shows, the proposed 3D CNN achieved a MAE of 28.9 days, a RMSE of 37.0 days, a MRE of 7.8%, a R of 0.983, and a R² of 0.967 in predicting brain age of children aged 0–2 years old, while a MAE of 110.0 days, a RMSE of 133.5 days, a MRE of 8.2%, a R of 0.883, and a R² of 0.780 were obtained in predicting

brain age of children aged 2–5 years old. It is found that the predictions for children under two years old are much better than that over two. Actually, this phenomenon is consistent with the understanding of clinical practice—that is, brain development under 2 years old is rapid and heterogeneous, while the brain over 2 years old develops relatively statically (1, 2). Slow development of the brain over 2 years old leads to low distinguishability and high prediction error.

Optimizing Model Parameters Can Improve Prediction Accuracy

Generally, the prediction performance is quite dependent on the structure of CNN and the hyper parameters. Thus, we optimized the 3D CNN structure and the hyper parameters with grid search for achieving the best performance on training set and reported the performance on the test set independently. Recently, some evidence reports that network depth is crucially

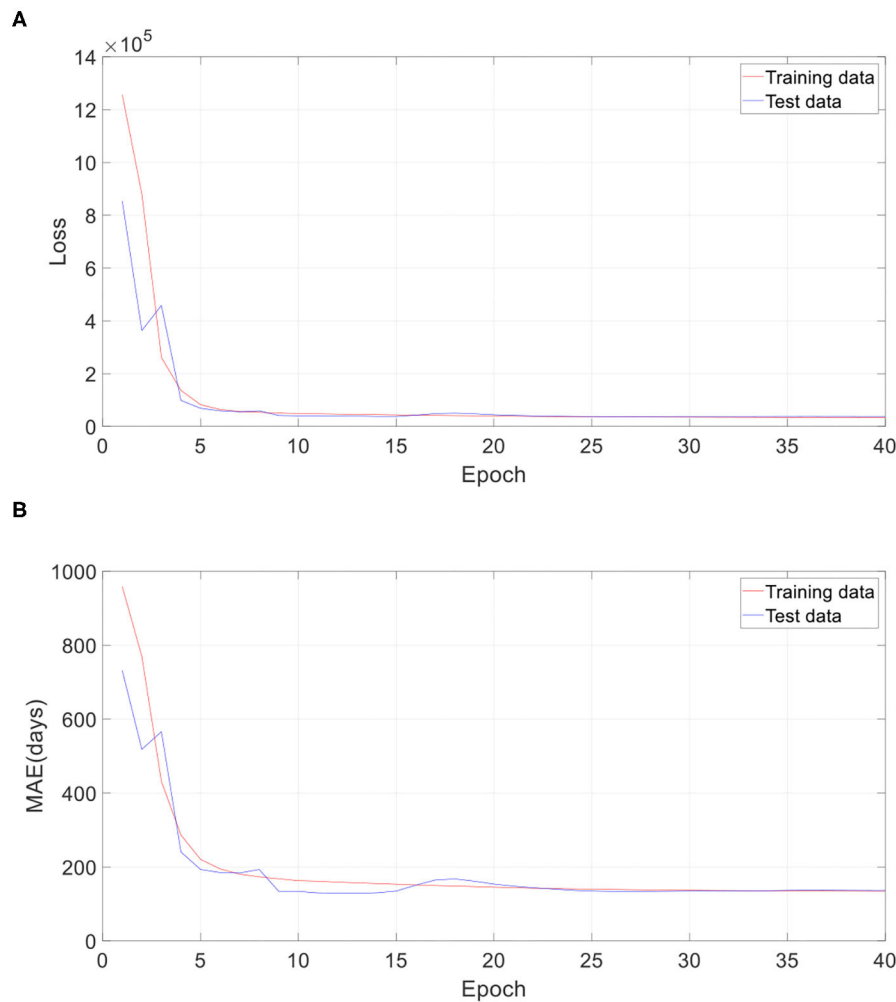


FIGURE 10 | Training performance of one run without batch normalization. **(A)** Loss against training epoch, and **(B)** MAE against training epoch. The loss and MAE are the average of all iterations in one epoch.

important for achieving remarkable prediction results (53, 54). Thus, we investigated the influence of different network depths on the prediction results, showing in **Table 4** and **Figure 9**. As we can see, the best performance was achieved by the 3D CNN containing seven convolution layers and three fully connected layers, not the deepest or shallowest network. Theoretically, the more convolution layers, the higher the extracted feature levels, and the more fully connected layers, the more complex the mapping function that can be fitted. However, too many neuron layers will produce redundant parameters, easily resulting in overfitting. Except for overfitting, a degradation problem may also occur when the deep network starts to converge: with the network depth increasing, accuracy gets saturated (55, 56). Therefore, in order to obtain the best performance, it is necessary to choose a network structure with the appropriate depth.

If the neural network is too deep, the gradient will become very small when it propagates back to the shallow layer, so that the parameters of the shallow layer cannot be updated or the

amplitude of the update is very small. This phenomenon called gradient dispersion will lead to the requirement of lower learning rate and careful parameter initialization. Batch normalization was developed to address the above problems (31). In this study, we firstly tried to set the initial learning rate of the 3D CNN without batch normalization the same as that of the proposed 3D CNN. However, it is found that the learning rate is too high, which leads to the failure of training the 3D CNN without batch normalization. With grid search and trial-and-error methods, we set the learning rate of the network to 0.000000008. This observation fully proves that the network without batch normalization requires more careful parameter setting. According to **Figures 5, 10**, the training loss of the 3D CNN without batch normalization is bigger than that of the proposed 3D CNN, indicating that the former fits the training data worse than the latter. Furthermore, as **Table 5** shows, we observed that batch normalization can greatly improve the prediction performance according to MAE, RMSE, R , and R^2 . Thus, batch normalization is strongly recommended for use in

TABLE 6 | Comparison of the proposed 3D CNN trained by different batch size and initial learning rate.

Batch size	Learning rate	MAE (days)	RMSE (days)	MRE (%)	<i>R</i>	<i>R</i> ²
16	0.0000008	72.3	96.2	8.3	0.985	0.971
32	0.0000008	74.1	97.4	9.2	0.985	0.971
64	0.0000008	67.6	96.1	8.2	0.985	0.971
64	0.0000012	69.9	97.1	8.6	0.985	0.971
64	0.0000004	75.0	104.8	9.3	0.983	0.966

TABLE 7 | Comparison of 2D CNN and our proposed 3D CNN.

	MAE (days)	RMSE (days)	MRE (%)	<i>R</i>	<i>R</i> ²
2D CNN	75.1	104.6	9.2	0.982	0.965
3D CNN	67.6	96.1	8.2	0.985	0.971

TABLE 8 | Comparison with state-of-the-art approaches.

Method	Subjects	Age (days)	MAE (days)
Feature-based developmental model (51)	92	8–590	72
HRtoF model (46)	50	14–797	32.1
The proposed 3D CNN	88	4–697	28.9
The proposed 3D CNN	220	4–1,820	67.6

3D CNN for predicting brain age using stacked 2D routine clinical brain MR images.

In terms of hyper parameters, we investigated the influence of batch size and learning rate on the performance of the 3D CNN. Basically, the larger the batch size, the more stable the gradient descent and the more accurate the direction. However, large batch size may cause the model to fall into local minimums and cannot come out because of the little noisiness. Small batch size may cause the data distribution to be too random to converge. Thus, the best batch size should be obtained by experiments for making the model converge to the global minimum as much as possible. In this paper, we set the batch size as 16, 32, and 64 for observing their effects on the predictions, showing in **Table 6**. It is found that the batch size of 64 achieved the best performance. The reason for not increasing the batch size is because the computer does not have enough computing memory, which is also the disadvantage of large size that cannot be ignored. Learning rate controls the convergence speed of model. When the learning rate is set too small, the convergence process becomes very slow and may make the model overfit. When the learning rate is set too large, the gradient may oscillate back and forth around the minimum value, and may not even converge. Thus, it is necessary to select the appropriate learning rate with grid search for achieving the best performance. As **Table 6** shows, the middle-sized learning rate yields the best predictions.

CONCLUSION

In this paper, we developed an end-to-end AI system based on 3D CNN for predicting the brain age of children aged 0 to 5 years old and achieved reliable and high performance with a MAE of 67.6 days, a RMSE of 96.1 days, a MRE of 8.2%, a *R* of 0.985, and a *R*² of 0.971. We found that the predictions for children under 2 years old are much better than those over two, which is also better than two state-of-the-art methods of predicting brain age of infants. The changes in the structure of the model have small effects on the prediction results, as do the changes in learning rate and batch size. The tricks of data augmentation and batch normalization have a significant impact on model performance. The proposed 3D CNN outperformed the 2D CNN having similar structure in prediction results.

In the future, we will collect more subjects for enhancing the performance of the model since CNN is a kind of data-driven method. Furthermore, we will enroll child patients with neurodevelopmental or mental disorders for validating the performance of the model in predicting the biological age of their brains.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material. Code Repository: <https://github.com/Captain-Hong/Brain-Age-Prediction-of-Children>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Children's Hospital of Nanjing Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ZF and MY collected the data. JH designed the algorithm. JH and Y-DZ preprocessed the data, designed the algorithm, and tested the model. S-HW, MY, and Y-DZ interpreted the results. JH and ZF drafted the work. MY and YZ gave guidance on experiment design. JH, ZF, and S-HW organized the literature. YS and AP substantively revise the manuscript. All authors gave critical comments and approved the submission.

FUNDING

This study was supported by Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Hope Foundation for Cancer Research, UK (RM60G0680); Medical Research Council Confidence in Concept (MRC CIC) Award, UK (MC_PC_17171); British Heart Foundation Accelerator Award, UK; Six Talent Peaks Project in Jiangsu Province, CN (WSN-192); China and Jiangsu commission of health, CN (LGY2019009).

REFERENCES

- Ducharme S, Albaugh MD, Nguyen TV, Hudziak JJ, Mateos-Pérez JM, Labbe A, et al. Trajectories of cortical thickness maturation in normal brain development — The importance of quality control procedures. *NeuroImage*. (2016) 125:267–79. doi: 10.1016/j.neuroimage.2015.10.010
- Watanabe M, Sakai O, Ozonoff A, Kussman S, Jara H. Age-related apparent diffusion coefficient changes in the normal brain. *Radiology*. (2013) 266:575–82. doi: 10.1148/radiol.12112420
- Welker K, Patton A. Assessment of normal myelination with magnetic resonance imaging. *Semin Neurol*. (2012) 32:015–28. doi: 10.1055/s-0032-1306382
- Gilmore JH, Lin W, Prastawa MW, Looney CB, Vetsa YSK, Knickmeyer RC, et al. Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *J Neurosci*. (2007) 27:1255–60. doi: 10.1523/JNEUROSCI.3339-06.2007
- Knickmeyer RC, Gouttard S, Kang C, Evans D, Wilber K, Smith JK, et al. A structural MRI study of human brain development from birth to 2 years. *J Neurosci*. (2008) 28:12176–82. doi: 10.1523/JNEUROSCI.3479-08.2008
- Gilmore JH, Shi F, Woolson SL, Knickmeyer RC, Short SJ, Lin W, et al. Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cerebral Cortex*. (2012) 22:2478–85. doi: 10.1093/cercor/bhr327
- Lyall AE, Shi F, Geng X, Woolson S, Li G, Wang L, et al. Dynamic development of regional cortical thickness and surface area in early childhood. *Cerebral Cortex*. (2015) 25:2204–12. doi: 10.1093/cercor/bhu027
- Li G, Nie J, Wang L, Shi F, Lin W, Gilmore JH, et al. Mapping region-specific longitudinal cortical surface expansion from birth to 2 years of age. *Cerebral Cortex*. (2013) 23:2724–33. doi: 10.1093/cercor/bhs265
- Wu M, Lu LH, Lowes A, Yang S, Passarotti AM, Zhou XJ, et al. Development of superficial white matter and its structural interplay with cortical gray matter in children and adolescents. *Human Brain Mapping*. (2014) 35:2806–16. doi: 10.1002/hbm.22368
- Barnea-Goraly N, Menon V, Eckert M, Tamm L, Bammner R, Karchemskiy A, et al. White matter development during childhood and adolescence: a cross-sectional diffusion tensor imaging study. *Cerebral Cortex*. (2005) 15:1848–54. doi: 10.1093/cercor/bhi062
- Fransson P, Skiöld B, Horsch B, Nordell A, Blennow M, Lagercrantz H, et al. Resting-state networks in the infant brain. *Proc Natl Acad Sci USA*. (2007) 104:15531–6. doi: 10.1073/pnas.0704380104
- Fransson P, Skiöld B, Engström M, Hallberg B, Mosskin M, Åden U, et al. Spontaneous brain activity in the newborn brain during natural sleep—an fMRI study in infants born at full term. *Pediatric Res*. (2009) 66:301–5. doi: 10.1203/PDR.0b013e3181b1bd84
- Gao W, Lin W, Grewen K, Gilmore JH. Functional connectivity of the infant human brain: plastic and modifiable. *Neuroscientist*. (2017) 23:169–84. doi: 10.1177/1073858416635986
- de Bie HM, Boersma M, Adriaanse S, Veltman DJ, Wink AM, Roosendaal SD, et al. Resting-state networks in awake five- to eight-year old children. *Human Brain Mapping*. (2012) 33:1189–201. doi: 10.1002/hbm.21280
- Baccouche M, Mamalet F, Wolf C, Garcia C, Baskurt A. *Sequential Deep Learning for Human Action Recognition*. Berlin; Heidelberg: Springer (2011). doi: 10.1007/978-3-642-25446-8_4
- Ji S, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern. Anal. Mach. Intell.* (2012) 35:221–31. doi: 10.1109/TPAMI.2012.59
- Wang L, Qiao Y, Tang X. Action recognition with trajectory-pooled deep-convolutional descriptors. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: Institute of Electrical and Electronics Engineers (IEEE). (2015). doi: 10.1109/CVPR.2015.7299059
- Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A. Learning deep features for scene recognition using places database. In: *Advances in Neural Information Processing Systems*. Montreal, CA: Neural Information Processing Systems Foundation, Inc. (NIPS) (2014)
- Li H, Lin Z, Shen X, Brandt J, Hua G. A convolutional neural network cascade for face detection. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: Institute of Electrical and Electronics Engineers (IEEE) (2015). doi: 10.1109/CVPR.2015.7299170
- Abdel-Hamid O, Mohamed A, Jiang H, Deng L, Penn G, Yu D. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions Audio Speech Language Processing*. (2014) 22:1533–45. doi: 10.1109/TASLP.2014.2339736
- Song Q, Zhao L, Luo X, Dou X. Using deep learning for classification of lung nodules on computed tomography images. *J Healthcare Eng*. (2017) 2017:8314740. doi: 10.1155/2017/8314740
- Gu Y, Lu X, Yang L, Zhang B, Yu D, Zhao Y, et al. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. *Comput Biol Med*. (2018) 103:220–31. doi: 10.1016/j.compbiomed.2018.10.011
- Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals Oncol*. (2018) 29:1836–42. doi: 10.1093/annonc/mdy166
- Hong J, Wang SH, Cheng H, Liu J. Classification of cerebral microbleeds based on fully-optimized convolutional neural network. *Multimedia Tools Applications*. (2020) 79:15151–69. doi: 10.1007/s11042-018-6862-z
- Hong J, Cheng H, Wang SH, Liu J. Improvement of cerebral microbleeds detection based on discriminative feature learning. *Fundamenta Informaticae*. (2019) 168:231–48. doi: 10.3233/FI-2019-1830
- Hong J, Cheng H, Zhang YD, Liu J. Detecting cerebral microbleeds with transfer learning. *Machine Vision Applications*. (2019) 30:1123–33. doi: 10.1007/s00138-019-01029-5
- Wang SH, Phillips P, Sui Y, Liu B, Yang M, Cheng H. Classification of alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling. *J Med Systems*. (2018) 42:85. doi: 10.1007/s10916-018-0932-7
- Sarraf S, Tofighi G. DeepAD: Alzheimer's disease classification via deep convolutional neural networks using MRI and fMRI. *BioRxiv*. (2016) 2016:070441. doi: 10.1101/070441
- Cole JH, Poudel RP, Tsagkrasoulis D, Caan MW, Steves C, Spector TD, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*. (2017) 163:115–24. doi: 10.1016/j.neuroimage.2017.07.059
- Wang J, Knol MJ, Tiulpin A, Dubost F, de Bruijne M, Vernooij MW, et al. Gray matter age prediction as a biomarker for risk of dementia. *Proc Natl Acad Sci USA*. (2019) 116:21213. doi: 10.1073/pnas.1902376116
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Francis B, David B, editors. *Proceedings of the 32nd International Conference on Machine Learning*. Lille: Proceedings of Machine Learning Research (PMLR) (2015). p. 448–56.
- Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Haifa: Omnipress. (2010). p. 807–814.
- Cui X, Goel V, Kingsbury B. Data augmentation for deep neural network acoustic modeling. *IEEE/ACM Transactions on Audio, Speech, Language Processing*. (2015) 23:1469–77. doi: 10.1109/TASLP.2015.2438544
- Jaitly N, Hinton GE. Vocal tract length perturbation (VTLP) improves speech recognition. In *Proceedings ICML Workshop on Deep Learning for Audio, Speech and Language*. Atlanta, GA: JMLR:W&CP. (2013).
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Lake Tahoe: Neural Information Processing Systems Foundation, Inc. (NIPS) (2012).
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. (1998) 86:2278–324. doi: 10.1109/5.726791
- Simard PY, Steinkraus D, Platt JC. Best practices for convolutional neural networks applied to visual document analysis. In: *Seventh International Conference on Document Analysis and Recognition 2003 Proceedings*. Edinburgh, UK: IEEE Computer Society. (2003).
- Sajedi H, Pardakhti N. Age prediction based on brain MRI image: a survey. *J Med Syst*. (2019) 43:279. doi: 10.1007/s10916-019-1401-7
- Cole JH, Leech R, Sharp DJ, Alzheimer's Disease Neuroimaging Initiative. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann Neurol*. (2015) 77:571–81. doi: 10.1002/ana.24367

40. Cole JH, Ritchie SJ, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle N, et al. Brain age predicts mortality. *Mol Psychiatry*. (2018) 23:1385–92. doi: 10.1038/mp.2017.62
41. Lancaster J, Lorenz R, Leech R, Cole JH. Bayesian optimization for neuroimaging pre-processing in brain age classification and prediction. *Front Aging Neurosci*. (2018) 10:28. doi: 10.3389/fnagi.2018.00028
42. Liem F, Varoquaux G, Kynast J, Beyer F, Kharabian Masouleh S, et al. Predicting brain-age from multimodal imaging data captures cognitive impairment. *NeuroImage*. (2017) 148:179–88. doi: 10.1016/j.neuroimage.2016.11.005
43. Huizinga W, Poot DHJ, Vernooij MW, Roshchupkin GV, Bron EE, Ikram MA, et al. A spatio-temporal reference model of the aging brain. *NeuroImage*. (2018) 169:11–22. doi: 10.1016/j.neuroimage.2017.10.040
44. Luders E, Cherbuin N, Gaser C. Estimating brain age using high-resolution pattern recognition: younger brains in long-term meditation practitioners. *NeuroImage*. (2016) 134:508–13. doi: 10.1016/j.neuroimage.2016.04.007
45. Wang B, Pham TD. MRI-based age prediction using hidden Markov models. *J Neurosci Methods*. (2011) 199:140–5. doi: 10.1016/j.jneumeth.2011.04.022
46. Hu D, Wu Z, Lin W, Li G, Shen D. Hierarchical rough-to-fine model for infant age prediction based on cortical features. *IEEE J Biomed Health Informatics*. (2019) 24:214–25. doi: 10.1109/JBHI.2019.2897020
47. Huang T, Chen H, Fujimoto R, Ito K, Wu K, Sato K, et al. Age estimation from brain MRI images using deep learning. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. (2017). doi: 10.1109/ISBI.2017.7950650
48. Lin L, Jin C, Fu Z, Zhang B, Bin G, Wu S. Predicting healthy older adult's brain age based on structural connectivity networks using artificial neural networks. *Computer Methods Prog Biomed*. (2016) 125:8–17. doi: 10.1016/j.cmpb.2015.11.012
49. Saha S, Pagnozzi A, George J, Colditz PB, Boyd R, Rose S, et al. *Investigating Brain Age Deviation in Preterm Infants: A Deep Learning Approach*. Cham: Springer International Publishing (2018). doi: 10.1007/978-3-030-00807-9_9
50. Liang H, Zhang F, Niu X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Human Brain Mapping*. (2019) 40:3143–52. doi: 10.1002/hbm.24588
51. Toews M, Wells WM, Zöllei L. A feature-based developmental model of the infant brain in structural MRI. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*. Berlin; Heidelberg: Springer. (2012). doi: 10.1007/978-3-642-33418-4_26
52. Dou Q, Chen H, Yu L, Zhao L, Qin J, Wang D, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Transactions Med Imag*. (2016) 35:1182–95. doi: 10.1109/TMI.2016.2528129
53. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Computer Sci*. (2014) arXiv:1409.1556.
54. Szegedy C, Wei L, Yangqing J, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: Institute of Electrical and Electronics Engineers (IEEE) (2015). doi: 10.1109/CVPR.2015.7298594
55. He K, Sun J. Convolutional neural networks at constrained time cost. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: Institute of Electrical and Electronics Engineers (IEEE) (2015). doi: 10.1109/CVPR.2015.7299173
56. Srivastava RK, Greff K, Schmidhuber J. Highway Networks. *Computer Sci*. (2015) arXiv:1505.00387.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hong, Feng, Wang, Peet, Zhang, Sun and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.