Robust Brain Age Estimation Based on sMRI via Nonlinear Age-Adaptive Ensemble Learning

Zhaonian Zhang, Richard Jiang, Ce Zhang, Bryan Williams, Ziping Jiang, Chang-Tsun Li, Paul Chazot, Nicola Pavese, Ahmed Bouridane, *Senior Member, IEEE*, and Azeddine Beghdadi, *Senior Member, IEEE*

Abstract—**Precise prediction on brain age is urgently needed by many biomedical areas including mental rehabilitation prognosis as well as various medicine or treatment trials. People began to realize that contrasting physical (real) age and predicted brain age can help to highlight brain issues and evaluate if patients' brains are healthy or not. Such age prediction is often challenging for single model-based prediction, while the conditions of brains vary drastically over age. In this work, we present an age-adaptive ensemble model that is based on the combination of four different machine learning algorithms, including a support vector machine (SVR), a convolutional neural network (CNN) model, and the popular GoogLeNet and ResNet deep networks. The ensemble model proposed here is nonlinearly adaptive, where age is taken as a key factor in the nonlinear combination of various single-algorithm-based independent models. In our age-adaptive ensemble method, the weights of each model are learned automatically as nonlinear functions over age instead of fixed values, while brain age estimation is based on such an age-adaptive integration of various single models. The quality of the model is quantified by the mean absolute errors (MAE) and spearman correlation between the predicted age and the actual age, with the least MAE and the highest Spearman correlation representing the highest accuracy in age prediction. By testing on the Predictive Analysis Challenge 2019 (PAC 2019) dataset, our novel ensemble model has achieved a MAE down to 3.19, which is a significantly increased accuracy in this brain age competition. If deployed in the real world, our novel ensemble model having an improved accuracy could potentially help doctors to identify the risk of brain diseases more accurately and quickly, thus helping pharmaceutical companies develop drugs or treatments precisely, and potential offer a new powerful tool for researchers in the field of brain science.**

Index Terms—**Brain age, biomarks, ensemble deep learning, mental healthcare, rehabilitation.**

I. Introduction

THE increasing aging population presents many acute challenges globally in the 21st century, with a profound impact on all aspects of life. Amongst them, brain function decline and neurodegenerative diseases in the aging population result in serious economic, medical, and societal issues to our society [1], [2]. In life science and biomedical domain, methods of predicting and assessing the risk of age-related neurodegeneration in the elderly and related treatments to reduce and reverse the process are one of the fundamental research topics [3]. Although brain aging is a natural process, there are individual differences in the changes of brain volume, cortical thickness, and white matter microstructure [4]–[6]. In addition, the degree of deviation in brain aging trajectory for a particular person from the average trajectory of healthy brain aging has been shown to reflect the individual's future risk of developing neurodegenerative diseases [7], [8]. Therefore, building models based on the characteristic patterns of brain aging within neuroimaging data and detecting the aging trajectories of individual brains offer a new perspective for studying brain aging differences [3].

The accurate prediction of brain age has not only critical scientific significance but also extensive clinical value [9]. Research has shown that along with the increased difference between the predicted brain age and the biological age, the risk of mortality or physical problems increases, together with the increased likelihood of early death [10]. Brain age estimation can diagnose patients with Alzheimer's disease [66], [67], psychiatric disorders [68], physical problems [69] and traumatic brain injuries [70] according to accelerated brain age. This method can also predict the conversion from mild cognitive impairment to Alzheimer's disease in the future [71]. This approach can not only diagnose disease, but also provides the basis for good living habits; for example, Steffener *et al.* [8] proved that high education and physical exercise can help make brain activity and keep young. Luders *et al.* [7] reported that the brains of people who meditate regularly are more

active than those of normal people of the same age. Also, the work Erus *et al.* [72] suggested that accelerated cognitive development is an important factor leading to accelerated brain development in young subjects. Cheng *et al.* [73] used a two-satge 3D convolutional network for brain age estimation. He *et al.* [74] utilized a global-local vision transformer to achieve a good accuracy on brain age estimation. Peng *et al.* [75] exploited a simple lightweight fully convolutional network to address the challenges on the brain age estimation. These single-method models have demonstrated their merits on this challenging topic.

Predicting brain age can also play a meaningful role in medical development, with clinical trials being an important part of clinical science [12], [13]. At present, many pharmaceutical companies across the world are committed to the research of medicine for the treatment of age-related diseases. However, the effect of these medications will not be obvious in the short term. Even experienced doctors cannot judge whether the drugs have played a role, so the curative effect may take several years to follow up. This problem makes it difficult for pharmaceutical companies to collect medical data, which restricts the research and development of aging diseases medicine [13]. Nevertheless, brain age estimation provides an alternative direction to address the problem in observing the effects of drugs, by the changes of predicted human brain age [13]. In recent years, deep learning has been the main approach for the estimation of brain age, as it can capture subtle changes in the brain through hierarchical feature representations in an end-to-end way [3]. Existing research has shown that the difference between predicted brain age and the participant's actual age is small for healthy people [3], [7], [10]. The development of deep learning in brain age estimation enables pharmaceutical companies to conduct follow-up investigations from the beginning of patients taking drugs, so as to know the effect of drugs in time and acquire patients' data at fast pace.

The process of brain condition detection by brain age estimation has basically two steps. Initially, we need to develop a model that can determine the biological age of a healthy person with this state of the brain based on brain neuroimaging data [3]. This model can be determined by training deep learning models on healthy samples. Subsequently, we would compare the predicted age and real age. If a sample's predicted brain age is older than his real age, it represents poor brain health. It is worth noting that the training data must be collected from healthy people, because the age predicted by the model would show the age of a person generated from healthy people under similar brain conditions. For example, if the training data contains samples with diseases, the predicted age will not represent the age that the patient should be at in this brain condition, consequently, the comparison between the predicted age and the true age is meaningless.

A frequently used method for brain age estimation is making a classification or regression for brain images [1], [14]. There exist several machine learning methods for this purpose. Previously, Huang *et al.* [41] applied CNNs in brain age estimation, and notably, Cole *et al.* [42] implemented a 3D CNN, which is trained on T1-weighted MRI, to predict brain age and achieved
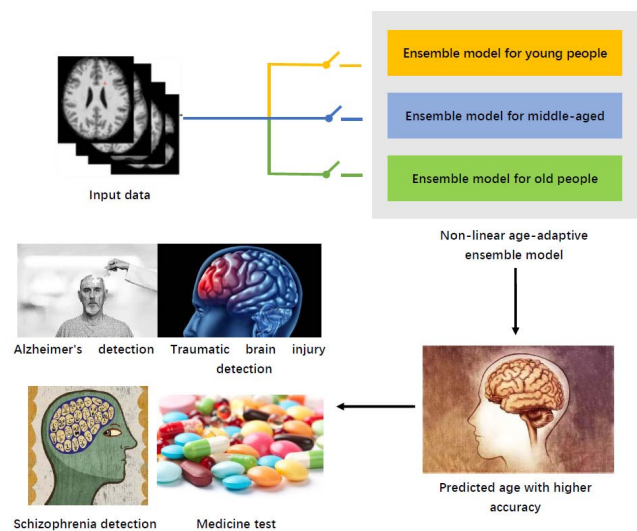

Fig. 1. The contributions of our AAE model.

promising results. Our initial motivation is to identify which of the various models for predicting brain age works best, and find the most suitable model for each age stage. Further, we aim to establish a novel ensemble model by combining different independent models together, and benchmark with single independent model on brain age prediction.

The major innovation of this work relates to a novel nonlinear age-adaptive ensemble model (nl-AAE), which is considered as a nonlinear function in the combination of multiple independent models. The age-adaptive ensemble model, with the advantages of multiple independent models, can be fully learned over the characteristics of the brain of each age group, thus achieving high accuracy of the predictions. Here, we have considered four different independent models, including a GoogLeNet, a ResNet, an SVR, and a self-designed CNN model. The nonlinear age-adaptive learning encoded in our ensemble model utilizes the changed weights of the constituent models based on the age of the sample. The combined model is adaptable to age changes nonlinearly and learns the brain characteristics over different ages.

We have tested our nl-AAE models using the PAC 2019 competition dataset, and benchmarked our models with four constituent algorithms. Such integrated model has great potential to provide a highly accurate measure of brain health for clinical trials of neuroprotective therapies, screening groups of people at-risk of poorer cognitive aging, and provide mechanistic insights into the downstream consequences of different aging-related diseases. Figure 1 shows the contributions of our AAE model, it has higher accuracy in predicting brain age compared to other classical methods, which can make it has better performance in Alzheimer's detection, traumatic brain injury detection, Schizophrenia detection, medicine testing and so on.

The remainder of the paper is organized as follows. Section II reviews the existing relevant work. Section III gives a preliminary overview on existing models. Section IV presents our proposed nonlinear age-adaptive ensemble model.

Section V shows the experimental results. Finally, Section VI concludes the whole paper.

II. RELATED WORK

Previously, brain age prediction is conducted using feature extraction with brain MRIs followed by a classification or regression analysis. However, useful information might be lost since the manually engineered features are not likely to explicitly describe the relevant information on brain age. To be specific, pre-processing the image subjectively requires additional assumptions at the various stages during the pre-processing pipeline. However, these assumptions can hardly be satisfied, which can result in a model error [16], [17]. Besides, extracting features manually is a time-consuming task. In practice, decisions should be made within a few minutes to avoid the delay of treatment in the application. The above issues are the main reasons why brain age prediction was not widely adopted.

The emergence of deep learning models provides a possibility to address those issues. Convolutional neural networks, which are widely adopted in image classification tasks, have shown great potential in visual feature extraction. The astonishing learning ability and automated decision-making pipeline of CNN models make it a perfect alternative for brain age estimation that can improve the efficiency for medical consultation, clinical diagnosis as well as treatment decision making [2], [18], [19]. At present, deep learning not only has successfully developed in the field of diagnosing schizophrenia [20], ADHD [21], autism [22], and Alzheimer's disease [23], but also helps to identify new biomarkers [24] and formulating new hypotheses [25].

Although deep learning has achieved success in biomedical fields, there are still several remaining challenges in terms of technology and practical applications [3], [26]–[28]. For example, deep neural networks require large sample sizes for fitting models, while neuroimaging datasets often have relatively smaller capacities [29], [30]. The data scarcity has restricted the ability to learn image features effectively, and the problem of overfitting can also appear. Compared with 2D neuroimaging data, 3D images require larger GPU memory, which means that successful models in 2D data are not necessarily feasible in 3D scenes (e.g., ImageNet classification [31], [32]). Besides, further improving the model accuracy is a long-term objective in deep learning research. Literature shows that deep learning models fail to achieve the best result for certain tasks [33], [34]. Another open question is how to choose the suitable complexity of the model. The no free lunch theorem suggests that task specified design is necessary for to achieve better result.

Ensemble modeling provides a solution in choosing the best predictive model in machine learning. An ensemble model combines the prediction from several models to make the final prediction, by which the overall performance of the model is increased [35]. Several strategies of combining the prediction from individual models were proposed, such as averaging, voting, to improve performance. As early as 1785, Marquis de Condorcet argued that if the probability of each independent
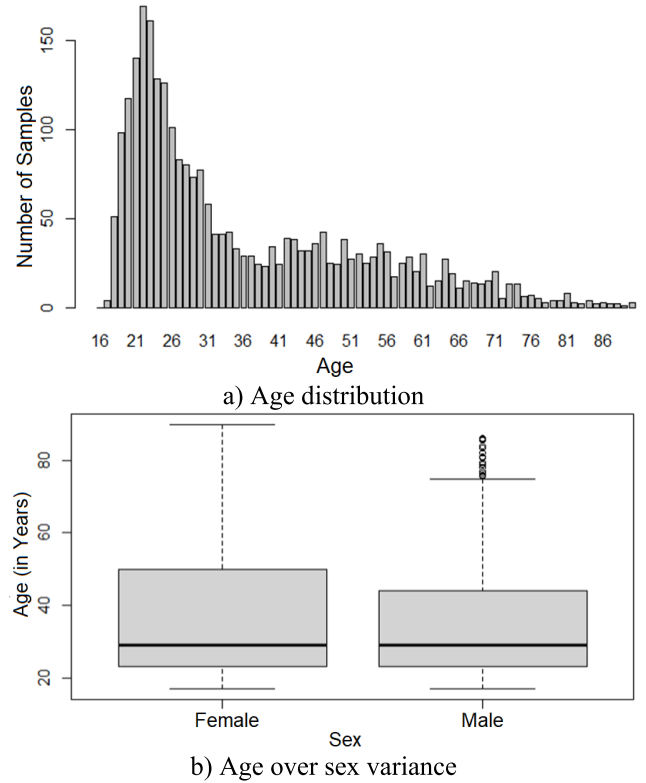

a) Age distribution

b) Age over sex variance

Fig. 2. Age and sex distribution of the MRI brain dataset.

voter being correct is above 0.5, then the addition of more voters increases the probability of the majority vote being correct [36], which is a strong evidence to show that ensemble models have better performance than individual models.

III. PRELIMINARIES ON BRAIN AGE ESTIMATION

A. Dataset

The dataset we used here is based on [3], including 2641 healthy individuals' brain sMRIs and information of samples such as their age and gender. The sample age ranged from 16 to 90 years old, the average age of samples is 35.8 years old, and the standard deviation of age is 16.2 years. Of the participants, 53% are females, and 47% are males. The average age of females is 37 years old, and the standard deviation of females' age is 17.2 years. The average age of males is 34.6 years old, and the standard deviation of males' age is 14.9 years. The age distribution of the data is shown in Figure 2. Here we remark that the dataset has an unbalanced distribution, with fewer data samples in the aged population and more data samples toward younger population. Cole *et al.* [42] shows the details of the samples.

B. Data Features

In our project, we use two different kinds of data as input for the models. One is Gray Matter and White Matter Maps, the other is Surface-Based Processing of Gray Matter.

The Gray Matter and White Matter Maps were distributed by the PAC organization. The pre-processing of nonlinear registration for the brain sMRIs used MNI152 space. Then, these images were segmented using DARTEL and SPM12 as different tissues, such as Gray Matter and White Matter so that

each tissue has a map and the map was smoothed by using a 4-mm kernel. For more details, please refer to [42]. This kind of data is used for the input of self-defined CNN, ResNet and GoogLenet in our project.

As for Surface-Based Processing of Gray Matter, we extract the vertex-wise measurements of cortical thickness and surface area based on the sMRIs by using FreeSurfer 6.0 [44]. As such. a vertex-wise feature of seven subcortical nuclei thickness and surface was also extracted by the ENIGMA-shape protocol [45], [46]. After these data-pre-processings, we get nearly 650,000 gray matter measurements per individual. This processing method was used by Baptiste Couvy-Duchesne *et al.* [47], and they proved that these processed data have a max association with age. This kind of data is used for the input of SVR in our project.

C. Basic Independent Models

Recent progress on deep neural networks [50]–[52] has greatly enlightened the applications of medical biometrics [53]–[55] for health diagnosis, particularly toward understanding how neurons in brain function [56] and dysfunction [57]. In this work, we aim to exploit deep learning techniques with ensemble approaches for our brain age estimation task.

Before we describe the architecture of the ensemble model, we first introduce the basic blocks of our proposed model. According to previous research, each of the models has remarkable performance in estimating brain age [3], [41], [42].

1) *Convolutional Neural Networks*: The CNN we have built was implemented using Keras with TensorFlow as backend. For the first 5 consecutive blocks, each of them consists of a $3 \times 3 \times 3$ 3D convolution layer, a Batch Norm layer, an ELU activation and a Max Pooling layer. As for the 6th block, it contains a dropout layer and the 7th block contains a fully connected layer. The input data is a 3D volume image of $121 \times 145 \times 121$ pixels, and the convolutional part of this model reduces this image to 128 feature maps of size $4 \times 5 \times 4$. The finally fully connected layer reduces the feature maps down to the numbers that stand for predicted ages.

   We train this model on two channels by using the concatenation of gray matter and white matter. The loss function is MAE, and the optimization machine is Adam. The learning rate is 0.001, the decay is 10^{-4}, $\beta1$ is 0.9 and $\beta2$ is 0.999.

2) *GoogLeNet (Inception V1)*: This structure is used for brain age estimation in [43]. It is composed of a stem network, two inception modules, a max-pooling layer, five inception modules (note that two of them are connected to an auxiliary regression each), a max-pooling layer, two inception modules, an average pooling layer, a dropout layer, and a fully connected layer. Compared to Google's Inception V1, it changes the softmax layer to a fully connected layer as the final layer so that this task becomes a regression task but not a classification task. The convolutional filter in this model consists of an input layer, a convolutional layer, a batch normalization layer,

a ReLU activation and an output layer. The stem network consists of an input layer, a convolutional filter, a max-pooling layer, two convolutional filters, a max-pooling layer and an output layer. In the inception modules, there is an input layer, seven convolutional filters, a max-pooling layer, a concatenation layer and an output layer. The auxiliary regression, which is used for mitigating the vanishing gradient problem, is composed of an input layer, an average pooling layer, a convolutional filter, a fully connected layer, a ReLU layer, a dropout layer, a fully connected layer and an output layer.

   The input data of this model is 3D maps of gray matter density with $121 \times 145 \times 121$ pixels, and the output is the predicted age. The loss function is MAE, and we use Adam as the model's optimization machine, the learning rate is 0.0001 and the batch size is 8.

3) *ResNet*: The parameters of the ResNet we built are similar to the above CNNs built by ourselves. The difference is that the ResNet includes residual blocks, while our self-built CNNs do not have these blocks. The ResNet consists of 5 residual blocks, each followed by a max pooling layer of kernel size $3 \times 3 \times 3$ and stride $2 \times 2 \times 2$, and one fully connected block. The residual block is a combination of layers which are repeated twice inside. This combination consists a 3D convolutional layer with stride $1 \times 1 \times 1$ and kernel size $3 \times 3 \times 3$, a batch renormalization layer, and an ELU activation function. It also adds the signal feeding into the residual block to the output of a layer close to the end of the block. The fully connected block is a multilayer perceptron which has one hidden layer. The input layer has $128 \times 4 \times 5 \times 4 = 10240$ neurons, there are 256 neurons that use an ELU activation function in the hidden layer (FC 1), and there is a single neuron in the output layer. A dropout layer, whose keep rate is 0.8, is employed following the hidden layer. And finally the output layer (FC 2) performs a linear regression on the hidden layer features.

   We use 3D maps of gray matter density as input data, and MAE as loss function. The model is optimized using Adam with a learning rate of 0.001. We set the decay is 10^{-4}, $\beta1$ is 0.9 and $\beta2$ is 0.999.

4) *SVR*: SVM is a classical machine learning model which construct a set of hyperplanes that separate the feature space. It was first used for the binary classification task, and then it was updated to the regression version called SVR, which can solve the regression tasks. In this work, we use SVR with a radial basis function kernel, the input data is Surface-Based Processing of Gray Matter, which has nearly 650,000 gray matter measurements per individual, and the output is the sample's predicted age. The implementation we used is package scikit-learn in Python. The number of epochs we set is over 300 to keep the models with the highest accuracy.

With the above four models, we will investigate these models over different age groups and establish an ensemble model based on these independent models.

IV. Proposed Age-Adaptive Ensemble Model

A. Fundamentals of Ensemble Learning

Ensemble learning completes learning tasks by constructing and combining multiple learners. It is also referred to as a multi-classifier system or committee-based learning. The general structure of an ensemble learning model is to generate a group of individual learners first, and then combine them with a certain fusion strategy. In general, the generalization performance of ensemble learning is better than the individual learners [58], [59].

There are six common types of ensemble learning models: Bayes optimal classifier, Boosting, Bootstrap aggregating (bagging), Bayesian model averaging (BMA), Bayesian model combination (BMC), and Stacking.

Bayes optimal classifier is based on Bayesian decision theory. It is an ensemble of all the hypotheses in the hypothesis space [60]. Until now, it is still a popular supervisor learning for the problem of classification.

Boosting is an algorithm that can boost weak learners to strong ones [61]. It first trains a base learner from the initial training set, and then adjust the distribution of training samples according to the performance of the base learner, so that the misclassified samples will receive more attention in the follow-up. This process is repeated until the number of base learners reaches the pre-specified value.

Bagging is the most famous representative of parallel type ensemble learning. Its principle is based on bootstrap sampling [62]. For a subset with capacity of m is constructed with sampling with replacement. For an ensample model with T base learners, T subsets are generated to train each of the learners. The prediction is then made by fuse the results from base learners. Random Forest is one of the most famous extended variants of Bagging.

BMA, BMC and Stacking represent different model combining strategies. BMA [63] uses the weighted average method to combine the models where the weight of each model is equal to the posterior probability of the model. BMC [64] is an algorithmic correction to BMA. Instead of sampling each model in the ensemble individually, it samples from the space of possible ensembles.

Stacking first trains the initial learner from the initial data set, and then generates a new data set for training the secondary learner. In this new data set, the output of the primary learner is used as the sample input feature, and the initial sample's label is still used as the sample label. In general, the secondary learner always uses the logistic regression model [65]. Stacking is usually provide better robustness than BMA and BMC, since BMA and BMC are sensitive to model approximation errors.

B. Proposed Nonlinear Age-Adaptive Ensemble Model

Through extensive experimentations, we found that the performance of all the models is influenced by the true age of samples (see part V). This indicates that some models are suitable for predicting young samples, and some models are suitable for older samples. In order to improve the prediction accuracy, we built a model named nonlinear age-adaptive

ensemble model. Different from the stacking strategy, our ensemble model can adjust weights of inside independent models according to the ground truth label.

The proposed framework is shown in Fig.3. First, we used different independent models as the initial learners. In our work, we employed four models: SVR, ResNet, GoogLeNet, and our own CNN. We used them to predict brain age, recorded the prediction results of these independent models, and then used these results as input values for the ensemble model.

Thereafter, we divided the sample into many groups by age, and in each group, there is an ensemble model which is combined by the independent models.

$$M = \sum \omega_i H_i \tag{1}$$

Here, M represents the prediction result of the ensemble model in a determined age group, H_i means the prediction result of the i-th independent model in this age group, and ω_i is the weight of the i-th independent model in this age group with $\sum \omega_i = 1$.

Our model adopts a novel method to decide the weights of independent models. In each age group, we set a loss function, its equation is shown below:

$$J(\omega) = \frac{1}{2}(H\omega - Y)^T(H\omega - Y) \tag{2}$$

H is an m×n-dimensional matrix, m is the number of samples, n is the number of independent models, ω is an n×1-dimensional vector, which is $(\omega_1, \omega_2, \ldots, \omega_n)^T$, ω_i means the weight of the i-th independent model, Y is a $m \times 1$-dimensional vector, which is $(y_1, y_2, \ldots, y_n)$, and y_i is the i-th sample's real age.

In this work, we tested two optimizers in minimizing the loss function. The gradient descent updates the weights by moving towards the steepest direction:

$$\omega = \omega - \alpha H^T(H\omega - Y) \tag{3}$$

where α is learning rate.

The ordinary least squares can also accomplish this task, it can be described as below:

$$\omega = (H^T H)^{-1} H^T Y \tag{4}$$

It is worth noting that the results from these two methods are the same after experiments.

For each age group, we then obtain a series of suitable weights for independent models, so that the model is able to adaptively combine the results from base models for different age groups. Formally, the age-adaptive model can be expressed as:

$$F(x) = \sum_{A \in age} \omega_A H(x, p_A) \tag{5}$$

Here, "age" represents the set of different age ranges, x is input data, ω_A is the value of ω at age A, and p_A is the parameters of independent models at age A.

The process of predicting the brain age of the sample is as follows. First, each independent model predicts the brain age of the sample, we record them as $(H_1, H_2, \ldots, H_n)$. Once done, we calculate the average of all predicted ages
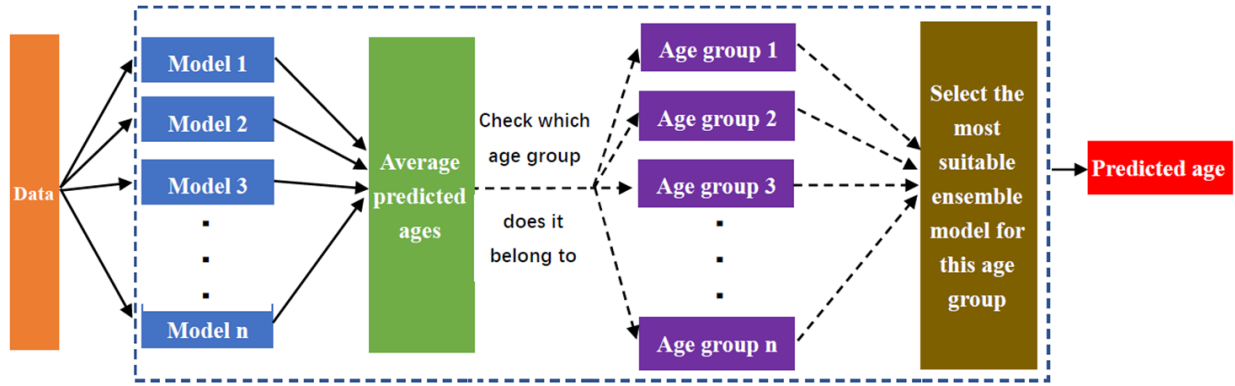
Fig. 3. The schematic view of our nonlinear age-adaptive ensemble method on brain age estimation.

and call it H_{ave}, $H_{ave} = mean(H_1 + H_2 + \ldots + H_n)$. Next, we check which age group H_{ave} belongs to. For each age group $G_i \in (G_1, G_2, \ldots, G_n)$, there is an ensemble model $M_i \in (M_1, M_2, \ldots, M_n)$ that combines the results of base models adaptively. In particular, if $H_{ave} \in G_i$, then we select M_i as the final ensemble model, and use it to predict the age of the sample. Figure 3 shows the estimation process of our ensemble model, and List I gives the list of the pseudocode of our method.

List 1 List of Pseudocode on Our Brain Age Estimation

Begin
    Input *brain sMRI* **as** *x*
    $H_1 = ResNet(x)$
    $H_2 = GoogLeNet(x)$
    $H_3 = CNN(x)$
    $H_4 = SVM(x)$
    $H_{ave} = mean(H_1 + H_2 + H_3 + H_4)$
    For $i = 1$ *to* n
      If $H_{ave} \in G_i$ **Then**
      Age – adaptive ensemble model $= M_i$
      End If
    End For
    Final result = Age – adaptive ensemble model (x)
    Output *Final result*
End

V. EXPERIMENTAL RESULTS

A. Experimental Results

The test method used is based on a 5-fold-cross-validation strategy, with the mean MAE and Spearman correlation between the predicted age and the actual age of them as final results.

The details of MAE for each model in a 5-fold-cross-validation are shown in Table I. Min means the minimum MAE, Max represents the maximum, and Mean is the average of 5 results in 5-fold-cross-validation while Std relates to the standard deviation of results, which represents the degree of dispersion of the results of the model. For the Std, Table I shows that the 6-layer CNN has the greatest value of 0.22,
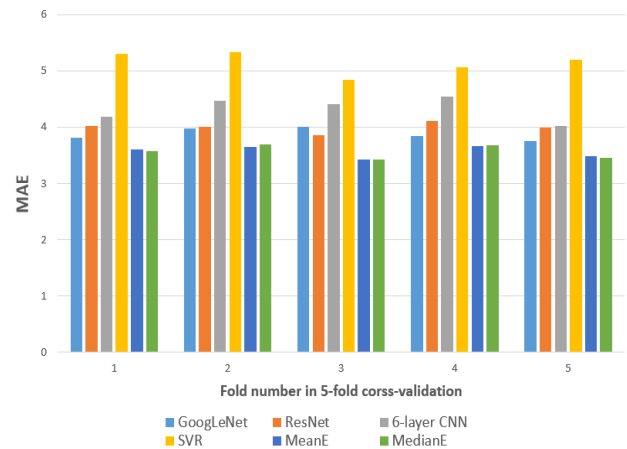

Fig. 4. Performance of each individual model.

followed by SVR with 0.21. For the ResNet architecture, Std value is 0.08, which is the minimum of all models. OE and MedianE have the same value of 0.13, nl-AAE-2 and nl-AAE-c are 0.12, GoogLeNet and nl-AAE-6 are 0.11, and MeanE is 0.1. These results suggest the results of 6-layer-CNN have the largest degree of dispersion. In contrast, the results of ResNet have the smallest degree of dispersion, which means the predictions of ResNet are more stable.

Our test results are shown in Figure 4 and Table II. In Table I, we first present the results of 4 independent models: SVR, 6-layer self-built CNN, ResNet and GoogLeNet. The achieved mean errors in term of years are 5.15, 4.33, 3.99 and 3.88 years in age, and the Spearman correlation between predicted age and real age are 0.83, 0.89, 0.88 and 0.89, respectively. We then combine the prediction of base learners together using the median and mean of their predictions. The result shows that the median based ensemble model has a larger mean error than the mean value based ensemble model. The possible reason is that a median-based ensemble model actually only chooses one model each time, and ignores other models that are not the median outputs.

Following these preliminary tests, we investigate the non-linear age-adaptive model. First, we used only one linearly approximated ensemble model to be applied to the data sample

TABLE I
THE DETAILS OF MAE FOR EACH MODEL IN 5-FOLD-CROSS-VALIDATION

	ResNet	6-layer CNN	GoogLeNet	SVR	MeanE	MedianE	OE	nl-AAE-2	nl-AAE-6	**nl-AAE-c**
Min	3.87	4.02	3.75	4.84	3.46	3.59	3.40	3.28	3.21	**2.98**
Max	4.11	4.54	4.00	5.34	3.71	3.92	3.69	3.65	3.55	**3.35**
Mean	3.99	4.33	3.88	5.15	3.57	3.72	3.52	3.42	3.39	**3.19**
Std	0.08	0.22	0.11	0.21	0.10	0.13	0.13	0.12	0.11	0.12

TABLE II
RESULTS OF ALL THE MODELS

	Model	Average MAE (years)	Spearman correlation
Single Model	SVR [44]	5.15	0.83
	6-layer CNN [43]	4.33	0.89
	ResNet [42]	3.99	0.88
	GoogLeNet [44]	3.88	0.89
Ensemble Method	MedianE [44]	3.72	0.9
	MeanE [44]	3.57	0.91
	OE [42]	3.52	0.91
	Our nl-AAE-2	3.45	0.89
	Our nl-AAE-6	3.39	0.95
	Our nl-AAE-c	**3.19**	**0.95**
Other researchers' methods	Seven algorithms combined ensemble model [44]	3.33	–
	Ensemble of shallow machine learning methods [49]	3.75	–
	Distributional Transformation [50]	4.58	0.93

of all ages. In other words, we first test the performance of ensemble model without age-adaptive. Compared with naive fusing strategies, the model performance is marginally improved with a MAE of 3.52 years, and the Spearman correlation of 0.91 (OE in Table II).

In our second experiment, we divided the prediction results of four independent models into two groups. The first group contains the samples over 40 years old while the other group contains the samples under 40 years old. We separately trained the models on two sets of data, and used the results of 4 base models as the input feature of secondary learner. By establishing two ensemble models for different groups and combining them into a non-linear ensemble model, a lower MAE with 3.45 years is achieved, but its Spearman correlation also drops to 0.89 (nl-AAE-2 in Table II).

Next, we divided the prediction results into six parts according to the actual age of the samples, which were 10-20 years old, 20-30 years old, 30-40 years old, 40-50 years old, 50-60 years old, and 60-90 years old, then we applied the same method to build the non-linear ensemble model (nl-AAE-6 in Table II). This time, the ensemble model's average MAE is 3.39 years, and its Spearman correlation improves to 0.95.

Finally, we divided the data more finely, taking all the same age sample as a group, but due to the small amount of data,

we can only use a simplified method, that is, for samples from 17 to 30 years old, we treated each age as a group. However, the size of data decreases as age increases. Therefore, for the 30 to 60-year-old samples, we took every 5 years old samples as a group. Likewise, we group the 60 to 70-year-old samples and the 70 to 90-year-old samples. We refer this finely split model as a "continuous" (or year-wise) model, namely nl-AAE-c in Table II. The ensemble model trained on this division provides the best performance with a MAE of 3.19 years and a Spearman correlation of 0.95.

Here, we present a comparison of our model with previous researchers' models which are also tested on the PAC 2019 dataset. Couvy-Duchesne et al. [43] built an ensemble model combined by 7 different algorithms, and its performance is attractive with a MAE of 3.33 years. Da Costa et al. [48] developed an ensemble of shallow machine learning methods (e.g., Support Vector Regression and Decision Tree-based regressors) with a MAE of 3.75 years. Soch [49] thought that distributional transformation (DT) can map the predicted values to the variable's distribution in the training data, which would improve decoding accuracy, and finally his model's performance is good with a MAE of 4.58 and spearman correlation between predicted age and the actual age of 0.93. These research results are attractive. By summarizing the experience of the previous researchers, we have developed the AAE method resulting in a marginal improvement compared to the previous research results.

The brain age gap represents the difference between predicted age and chronological age. Figure 5 shows the brain age gap as functions of the chronological age using 7 different machine-learning methods. The slope of the line in Figure 5 indicates how much the prediction accuracy of the model is affected by increasing the age. As such, the prediction accuracy of AAE has the lowest influence by aging, and the prediction accuracy of SVR is most affected by aging.

The above experiment provides the following inspirations: 1) deep neural networks are in-general better than SVR; 2) all ensemble models have lower errors than discrete models; 3) our age-adaptive models have better performance than non-adaptive ensemble models; 4) finer the age-based division, lower the error can be achieved by the nl-AAE-c model.

B. Investigation on Age-Sensitivity per Models

The age-sensitivity shows the trend of MAE for each model according to the age. Therefore, in this section, we investigate the age-sensitivity of the models we built previously. Figure 6 shows the results suggesting that all the models are good at predicting young people but not old people where the MAE
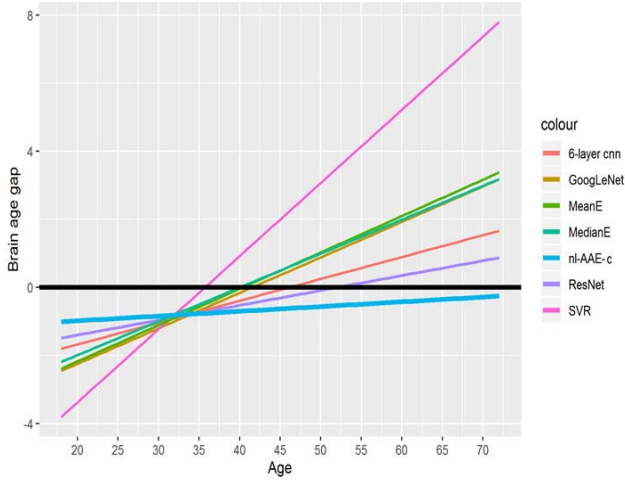
Fig. 5. The brain age gap and MAE as functions of the chronological age using 7 different machine-learning methods, the horizontal black line represents 0 brain age gap.
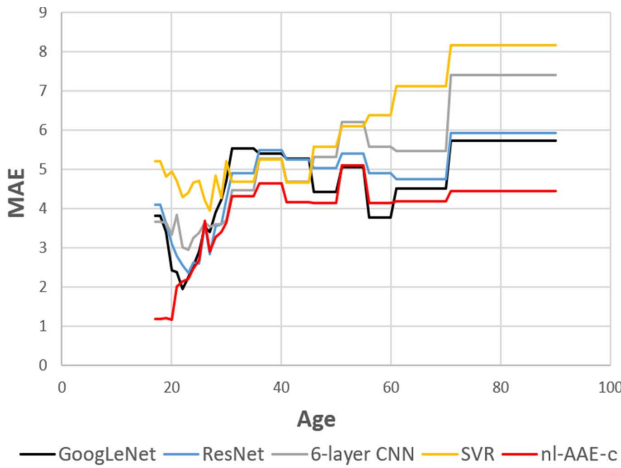

Fig. 6. Age-sensitivity of models.

increases with the age of the sample. As for independent models, GoogLeNet and ResNet are more sensitive for age, their MAE has a significant increase in the age of 20 to 30, but SVR does not change drastically as a whole. Besides, GoogLeNet has the best performance for middle-aged people, and the MAE of all models has a significant change when the sample age is 70 years old.

The age-sensitivity of nonlinear age-adaptive ensemble model is similar to that of the independent models but more stable. Obviously, as the sample age increases, the MAE becomes larger and the model's performance gets worse and worse. When the samples' age is 50 to 60, the model's performance is the worst, with MAE exceeding 5. For a machine learning perspective, this is due to the lack of older samples that resulted in an insufficient training of the model. From a medical perspective, we believe that this is due to the fact that the differences between the brains of different people will become larger as age grows, which in turn will increase the difficulty of prediction. On the other hand, brain differences between young people are not that big. Therefore,
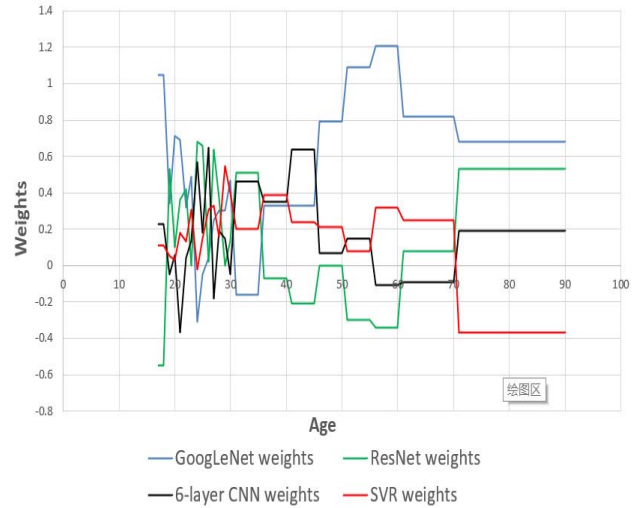

Fig. 7. Individual models' weights in nl-AAE-c.

when training the model in the future, we can increase the proportion of young people's data to improve the accuracy of the model.

C. Learning the Model Weights

Figure 7 shows the change of each independent model's weights in the AAE according to the age. Through this figure, we can know the different importance of each independent model in the AAE at different ages.

The SVR's weights are relatively average for a sample between 25 and 70 years old. But for young samples from 10 to 25 years old, SVR does not perform well, which shows that SVR is not suitable for predicting the age of young people.

The CNN model plays an important role in the prediction results when the sample age ranges from 20 to 50 years old. CNN receives a high weight in the prediction of the data of young samples, while for the data of old samples, CNN has less contribution to the ensembled result.

GoogLeNet model is important in the AAE among elder groups. Although it has average weights in the 20-40 age groups, its prediction significantly determines the results of the ensemble models in the elder groups, especially for middle-aged. This shows that the function of GoogLeNet is powerful, and it is suitable to predict the age of middle-aged and elderly samples.

ResNet model has basically maintained high weights for samples aged 10-35 years old, which shows that it is suitable for predicting the age of young samples. In the data of middle-aged and elderly people aged 35 to 70, it has average performance therefore it does not have a large weight. But for data over 70 years old, it has a good performance, which means that ResNet is suitable for inferring the age of people over 70 years old.

VI. DISCUSSION

A. Discussion of AI Models

In this work, we have proposed a nonlinear age-adaptive ensemble (nl-AAE) model for brain age estimation with better

results than other benchmark models. The most important contribution of this work is the age adaptive fusing strategy which significantly improves the performance of the ensemble model. To the best of our knowledge, no previous literatures documented this strategy. The characteristic of AAE is that it not only combines the advantages of multiple models, but the weights of independent models also change with age. This makes the AAE become a dynamic model, where the prediction results are more accurate given BrainAge projection is sensitive to age change.

Although our model yields state-of-art prediction results, there are still several issues that can be improved in future works. First, more advanced models can be introduced as initial models. For example, as for GoogLeNet, we used the Inception V1 version, now it has already had Inception V4 versions.

Second, MAE is generally used to evaluate the performance of the model. However, MAE is affected by age distribution and the number of objects in the training set, so MAE of different data sets cannot be directly compared. The lower the physiological age of the general object, the smaller the brain difference and the smaller the MAE value of the same age individuals. Actually for adolescents, the MAE value of the monomodular prediction model is 1 to 2 years, and that of the multimodal prediction model is about 1 year. However, for individuals of all ages or middle-aged and old age, the MAE value of the prediction model can only reach 4-5 years in general. Meanwhile, the larger the overall age span of the object is, the larger the evaluation index MAE will be. Therefore, the comparison between models can be made with various factors.

Third, we can try more ensemble methods, and compare their performance for brain age estimation in our future work. The approach we have used in this work is to re-weight the results of multiple independent models by integration and utilizing a nonlinear function which can have many types. For example, one can select only the best performing independent models as the ensemble model for each age group. In addition, we can also change the method of re-weighting. Using a multiple layer perceptron to combine the results of independent models is an interesting idea, although its training speed may be a little slower, its prediction accuracy is worth investigating.

Fourth, the gender is always an important variable in many experiments, based on the subtle differences between men's and women's brains. Therefore, the gender may be a key factor in improving model prediction accuracy. In our future work, we will investigate the influence of gender on brain age estimation and consider gender as a factor in our experiments.

Brain age prediction is a burgeoning research field that is developing rapidly. Brain age prediction models based on neuroimaging and their applications are increasing day by day. A growing number of researchers are using brain age analysis to explore brain aging in the course of health and disease, and many new and promising avenues of research are emerging. From the perspective of image modes, various image modes have their advantages and disadvantages, and the fusion of information from multiple modes is more likely to further improve the performance of the model. In addition, with the improvement of the architecture of the convolutional neural network and the appearance of the image data set of big data, we believe that the performance of the future model is likely to be further improved. The key to future model development is to continuously improve the accuracy of the model while improving the generalization ability of the model for new data. The ultimate goal of this field is to build a brain age model based on large image sets that completes that can be applied to provide accurate personalized cloud diagnosis services.

B. Discussion of Medical Findings

From our research, we also have some interesting findings from medical perspectives. First, we found that the performance of constituent models decreased with the age of the sample, which implies that young people have their brains in similar conditions. In addition, we have found that with an increase of age, the risk of people suffering from brain diseases increases, too. The differences between the brains are getting larger, and the accuracy of the model's prediction begins to decline. However, these results may be influenced by biological reasons, sample size or problems with the model, or a combination of both, so in the future, we still need to find a fairer way to prove these findings.

From the results of our experiments, we believe that changes in the brain can be divided into 4 stages, namely 0-30 years old, 30-50 years old, 50-70 years old and 70-80 years old. The criterion for classification is whether there is a significant change in the model's performance in predicting age. From Figure 5, we notice that during the period of 0-30 years old, the human brain undergoes a significant process of change. Speaking at a meeting of the Academy of Medical Sciences in Oxford in the UK, researchers explained that our brains slowly transition to adulthood, which is finally reached in our 30s. After the age of 30, the brain's working memory capacity begins to slowly decline [39], which is in agreement with our research findings. At the age of 30 to 50, the brain changes little, but there will be a significant change around the age of 50. Research in the British Medical Journal [40] also shows that in a group of people who were first tested on various mental abilities when they were 45–49 years old, reasoning skills declined by 3.6 percent over 10 years. At the age of 50-70, the brain does not change much, but after the age of 70, the brain will have the last significant change. Peter Jones's research also shows that the overall volume of the brain begins to shrink when we're in our 30s or 40s, with the rate of shrinkage increasing around age 60-70 [39], the results of our experiments can also be evidence of it.

Exercise, reading, meditation and other similar behaviors are good methods to prevent brain disease [7], [8]. People who exercise, meditate regularly, and those with higher education levels have lower predicted brain age than their peers, which shows that their brains are more active and the risk of brain diseases is lower. A study reported in [37] analyzed samples over the age of 50 and found that people who do not exercise or who exercise little have their brains about 5-10 years older than those who exercise regularly. Another recent research

found that stem cells in the brain's hypothalamus likely control how fast aging occurs in the body [38]. Specifically, the number of hypothalamic neural stem cells naturally declines over the life of the animal, and this decline accelerates aging. Researchers injected hypothalamic stem cells into the brains of normal old and middle-aged mice, whose stem cells had been destroyed, the measures of aging were slowed or reversed. This is an exciting discovery, which will be an important step in slowing down aging and treating brain diseases. The brain age prediction model in this article is sensitive to changes in the brain, and we believe it can be a useful tool for detecting medicine performance.

VII. Conclusion

In this paper, we proposed a nonlinear age-adaptive ensemble method for brain age estimation from MRI images. From our experiments, we clearly show that ensemble models can in general achieve lower errors than discrete models, and our nonlinear age-adaptive ensemble models are consistently better than age-agnostic ensemble models. Among discrete models, GoogLeNet basically has a good performance on data of all age groups, especially for middle-aged and old samples. With the significantly increased accuracy on brain age estimation, our nonlinear age-adaptive ensemble models can potentially help doctors to identify the risk of brain diseases more accurately and efficiently, help pharmaceutical companies develop drugs or treatments more precisely, and provide a powerful tool for researchers in the field of brain science.

References

[1] A. Reeve, E. Simcox, and D. Turnbull, "Ageing and Parkinson's disease: Why is advancing age the biggest risk factor?" *Ageing Res. Rev.*, vol. 14, pp. 19–30, Mar. 2014.

[2] R. Jiang, P. Chazot, N. Pavese, D. Crookes, A. Bouridane, and M. E. Celebi, "Private facial prediagnosis as an edge service for Parkinson's DBS treatment valuation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2703–2713, Jun. 2022.

[3] J. H. Cole and K. Franke, "Predicting age using neuroimaging: Innovative brain ageing biomarkers," *Trends Neurosci.*, vol. 40, no. 12, pp. 681–690, Dec. 2017.

[4] A. R. Groves *et al.*, "Benefits of multi-modal fusion analysis on a large-scale dataset: Life-span patterns of inter-subject variability in cortical morphometry and white matter microstructure," *NeuroImage*, vol. 63, no. 1, pp. 365–380, Oct. 2012, doi: 10.1016/j.neuroimage.2012.06.038.

[5] A. B. Storsve *et al.*, "Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: Regions of accelerating and decelerating change," *J. Neurosci.*, vol. 34, no. 25, pp. 8488–8498, Jun. 2014.

[6] A. M. Fjell *et al.*, "One-year brain atrophy evident in healthy aging," *J. Neurosci.*, vol. 29, no. 48, pp. 15223–15231, Dec. 2009, doi: 10.1523/JNEUROSCI.3252-09.2009.

[7] E. Luders, N. Cherbuin, and C. Gaser, "Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners," *NeuroImage*, vol. 134, pp. 508–513, Jul. 2016, doi: 10.1016/j.neuroimage.2016.04.007.

[8] J. Steffener, C. Habeck, D. O'Shea, Q. Razlighi, L. Bherer, and Y. Stern, "Differences between chronological and brain age are related to education and self-reported physical activity," *Neurobiol. Aging*, vol. 40, pp. 138–144, Apr. 2016, doi: 10.1016/j.neurobiolaging.2016.01.014.

[9] F. X. Castellanos, A. Di Martino, R. C. Craddock, A. D. Mehta, and M. P. Milham, "Clinical applications of the functional connectome," *NeuroImage*, vol. 80, pp. 527–540, Oct. 2013, doi: 10.1016/j.neuroimage.2013.04.083.

[10] J. H. Cole *et al.*, "Brain age predicts mortality," *Mol. Psychiatry*, vol. 23, pp. 1385–1392, May 2018, doi: 10.1038/mp.2017.62.

[11] B. Bulik-Sullivan *et al.*, "Analysis of shared heritability in common disorders of the brain," *Science*, vol. 360, no. 6395, 2018, Art. no. eaap8757.

[12] D. Bzdok, G. Varoquaux, and E. W. Steyerberg, "Prediction, not association, paves the road to precision medicine," *J. Amer. Med. Assoc. Psychiatry*, vol. 78, no. 2, p. 127, Feb. 2021, doi: 10.1001/jamapsychiatry.2020.2549.

[13] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: A practical introduction," *BMC Med. Res. Methodol.*, vol. 19, p. 64, Mar. 2019, doi: 10.1186/s12874-019-0681-4.

[14] E. E. Bron *et al.*, "Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI: The CADDementia challenge," *NeuroImage*, vol. 111, pp. 562–579, May 2015, doi: 10.1016/j.neuroimage.2015.01.048.

[15] A. Abbott, "Dementia: A problem for our age," *Nature*, vol. 475, no. 7355, pp. S2–S4, Jul. 2011.

[16] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008, doi: 10.1016/j.media.2007.06.004.

[17] X. Liu, M. Niethammer, R. Kwitt, N. Singh, M. McCormick, and S. Aylward, "Low-rank atlas image analyses in the presence of pathologies," *IEEE Trans. Med. Imag.*, vol. 34, no. 12, pp. 2583–2591, Dec. 2015, doi: 10.1109/TMI.2015.2448556.

[18] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, pp. 1342–1350, 2018, doi: 10.1038/s41591-018-0107-6.

[19] S. A. A. Kohl *et al.*, "A probabilistic U-Net for segmentation of ambiguous images," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 6965–6975.

[20] L. L. Zeng *et al.*, "Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI," *EBioMedicine*, vol. 30, pp. 74–85, Apr. 2018, doi: 10.1016/j.ebiom.2018.03.017.

[21] L. Zou, J. Zheng, C. Miao, M. J. Mckeown, and Z. J. Wang, "3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI," *IEEE Access*, vol. 5, pp. 23626–23636, 2017, doi: 10.1109/ACCESS.2017.2762703.

[22] R. M. Thomas, S. Gallo, L. Cerliani, P. Zhutovsky, A. El-Gazzar, and G. van Wingen, "Classifying autism spectrum disorder using the temporal statistics of resting-state functional MRI data with 3D convolutional neural networks," *Frontiers Psychiatry*, vol. 11, p. 440, May 2020, doi: 10.3389/fpsyt.2020.00440.

[23] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Med. Image Anal.*, vol. 43, pp. 157–168, Jan. 2018, doi: 10.1016/j.media.2017.10.005.

[24] M. D. Rosenberg, B. J. Casey, and A. J. Holmes, "Prediction complements explanation in understanding the developing brain," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, Dec. 2018, doi: 10.1038/s41467-018-02887-9.

[25] G. Shmueli, "To explain or to predict?" *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, 2010, doi: 10.1214/10-STS330.

[26] J. Kawahara *et al.*, "BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, Feb. 2017, doi: 10.1016/j.neuroimage.2016.09.046.

[27] S. Arslan, S. I. Ktena, B. Glocker, and D. Rueckert, "Graph saliency maps through spectral convolutional networks: Application to sex classification with brain connectivity," in *Graphs in Biomedical Image Analysis and Integrating Medical Imaging and Non-Imaging Modalities* (Lecture Notes in Computer Science), vol. 11044. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-030-00689-1_1.

[28] C. F. Baumgartner, L. M. Koch, K. C. Tezcan, J. X. Ang, and E. Konukoglu, "Visual feature attribution using Wasserstein GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8309–8319, doi: 10.1109/CVPR.2018.00867.

[29] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[30] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, "Transfusion: Understanding transfer learning for medical imaging," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32. Red Hook, NY, USA: Curran Associates, 2019, pp. 3347–3357.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

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

[33] T. He *et al.*, "Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics," *NeuroImage*, vol. 206, pp. 116–276, 2020, doi: 10.1016/j.neuroimage.2019.116276.

[34] M. A. Schulz *et al.*, "Different scaling of linear models and deep learning in U.K. Biobank brain images versus machine-learning datasets," *Nature Commun.*, vol. 11, p. 4238, 2020, doi: 10.1038/s41467-020-18037-z.

[35] A. Veit, M. J. Wilber, and S. Belongie, "Residual networks behave like ensembles of relatively shallow networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 550–558.

[36] M. D. Condorcet, *Essay on the Application of Analysis to the Probability of Majority Decisions*. Paris, France: Imprimerie Royale, 1785.

[37] J. Z. Willey *et al.*, "Leisure-time physical activity associates with cognitive decline, The northern Manhattan study," *Neurology*, vol. 86, no. 20, pp. 1897–1903, May 2016, doi: 10.1212/WNL.0000000000002582.

[38] Y. Zhang *et al.*, "Hypothalamic stem cells control ageing speed partly through exosomal miRNAs," *Nature*, vol. 548, no. 7665, pp. 52–57, Aug. 2017, doi: 10.1038/nature23282.

[39] J. Y. Guo *et al.*, "Brain structural changes in women and men during midlife," *Neurosci. Lett.*, vol. 615, pp. 107–112, Feb. 2016.

[40] F. Grodstein, "How early can cognitive decline be detected?" *BMJ*, vol. 344, p. d765, Jan. 2011.

[41] T.-W. Huang, "Age estimation from brain MRI images using deep learning," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 849–852.

[42] J. H. Cole *et al.*, "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.

[43] B. Couvy-Duchesne *et al.*, "Ensemble learning of convolutional neural network, support vector machine, and best linear unbiased predictor for brain age prediction: ARAMIS contribution to the predictive analytics competition 2019 challenge," *Frontiers Psychiatry*, vol. 11, Dec. 2020, Art. no. 593336.

[44] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, Aug. 2012, doi: 10.1016/j.neuroimage.2012.01.021.

[45] B. A. Gutman, S. K. Madsen, A. W. Toga, and P. M. Thompson, "A family of fast spherical registration algorithms for cortical shapes," in *Multimodal Brain Image Analysis* (Lecture Notes in Computer Science), vol. 8159, L. Shen, T. Liu, P. T. Yap, H. Huang, D. Shen, and C. F. Westin, Eds. Cham, Switzerland: Springer, 2013, doi: 10.1007/978-3-319-02126-3_24.

[46] B. A. Gutman, Y. Wang, P. Rajagopalan, A. W. Toga, and P. M. Thompson, "Shape matching with medial curves and 1-D groupwise registration," in *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI)*, May 2012, pp. 716–719, doi: 10.1109/ISBI.2012.6235648.

[47] B. Couvy-Duchesne *et al.*, "A unified framework for association and prediction from vertex-wise grey-matter structure," *Hum. Brain Mapping*, vol. 41, no. 14, pp. 4062–4076, 2020, doi: 10.1002/hbm.25109.

[48] P. F. Da Costa, J. Dafflon, and W. H. L. Pinaya, "Brain-age prediction using shallow machine learning: Predictive analytics competition 2019," *Frontiers Psychiatry*, vol. 11, Dec. 2020, Art. no. 604478.

[49] J. Soch, "Distributional transformation improves decoding accuracy when predicting chronological age from structural MRI," *Frontiers Psychiatry*, vol. 11, Dec. 2020, Art. no. 604268.

[50] Z. Jiang, P. L. Chazot, M. E. Celebi, D. Crookes, and R. Jiang, "Social behavioral phenotyping of Drosophila with a 2D–3D hybrid CNN framework," *IEEE Access*, vol. 7, pp. 67972–67982, 2019.

[51] C.-Y. Chiang, C. Barnes, P. Angelov, and R. Jiang, "Deep learning-based automated forest health diagnosis from aerial images," *IEEE Access*, vol. 8, pp. 144064–144076, 2020.

[52] F. Mehboob *et al.*, "Towards robust diagnosis of COVID-19 using vision self-attention transformer," *Nature Sci. Rep.*, vol. 12, May 2022, Art. no. 8922.

[53] R. Jiang, L. Zhang, H.-L. Wei, D. Crookes, and P. Chazot, Eds., *Recent Advances in AI-Enabled Automated Medical Diagnosis*, 1st ed. Boca Raton, FL, USA: CRC Press, 2022, doi: 10.1201/9781003176121.

[54] R. Jiang, C. Li, D. Crooke, W. Meng, and C. Rosenberger, Eds., *Deep Biometrics*. Cham, Switzerland: Springer, 2019, pp. 71–970, doi: 10.1007/978-3-030-32583-1.

[55] R. Jiang, S. Al-Madeed, A. Bouridane, D. Crookes, and A. Beghdadi, Eds., *Biometric Security and Privacy: Opportunities & Challenges in the Big Data Era*. Cham, Switzerland: Springer, 2016, doi: 10.1007/978-3-319-47301-7.

[56] R. Jiang and D. Crookes, "Shallow unorganized neural networks using smart neuron model for visual perception," *IEEE Access*, vol. 7, pp. 152701–152714, 2019.

[57] Z. Jiang, Y. Wang, C.-T. Li, P. Angelov, and R. Jiang, "Delve into activations: Towards understanding dying neuron," *IEEE Trans. Artif. Intell.*, early access, Jun. 9, 2022, doi: 10.1109/TAI.2022.3180272.

[58] L. Kuncheva and C. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, pp. 181–207, May 2003.

[59] P. Sollich and A. Krogh, "Learning with ensembles: How overfitting can be useful," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 8, 1996, pp. 190–196.

[60] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, no. 2, pp. 131–163, Nov. 1997, doi: 10.1023/A:1007465528199.

[61] Z.-H. Zhou, "The term boosting refers to a family of algorithms that are able to convert weak learners to strong learners," in *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012, p. 23

[62] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/BF00058655.

[63] C. Fraley, A. E. Raftery, T. Gneiting, and J. M. Sloughter, "ensembleBMA: An R package for probabilistic forecasting using ensemble and Bayesian model averaging," Dept. Statist., Univ. Washington, Tech. Rep. 516R, 2007.

[64] K. Monteith, J. L. Carroll, K. Seppi, and T. Martinez, "Turning Bayesian model averaging into Bayesian model combination," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2011, pp. 2657–2663.

[65] L. Breiman, "Stacked regressions," *Mach. Learn.*, vol. 24, no. 1, pp. 49–64, Jul. 1996, doi: 10.1007/BF00117832.

[66] K. Franke, G. Ziegler, S. Klöppel, and C. Gaser, "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *NeuroImage*, vol. 50, no. 3, pp. 883–892, Apr. 2010. [Online]. Available: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20070949&retmode=ref&cmd=prlinks, doi: 10.1016/j.neuroimage.2010.01.005.

[67] K. Franke and C. Gaser, "Longitudinal changes in individual BrainAGE in healthy aging, mild cognitive impairment, and Alzheimer's disease," *J. Gerontopsychol. Geriatric Psychiatry*, vol. 25, no. 4, pp. 235–245, 2012. [Online]. Available: http://www.psycontent.com/index/79311259G2307246.pdf, doi: 10.1024/1662-9647/a000074.

[68] N. Koutsouleris *et al.*, "Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders," *Schizophrenia Bull.*, vol. 40, no. 5, pp. 1140–1153, 2014. [Online]. Available: http://schizophreniabulletin.oxfordjournals.org/content/early/2013/10/11/schbul.sbt142.full, doi: 10.1093/schbul/sbt142.

[69] K. Franke, M. Ristow, and C. Gaser, "Gender-specific impact of personal health parameters on individual brain aging in cognitively unimpaired elderly subjects," *Frontiers Aging Neurosci.*, vol. 6, p. 94, May 2014. [Online]. Available: http://journal.frontiersin.org/article/10.3389/fnagi.2014.00094/abstract

[70] J. H. Cole, R. Leech, and D. J. Sharp, "Prediction of brain age suggests accelerated atrophy after traumatic brain injury," *Ann. Neurol.*, vol. 77, no. 4, pp. 571–581, Apr. 2015. [Online]. Available: http://doi.wiley.com/10.1002/ana.24367

[71] 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, Jun. 2013, Art. no. e67346. [Online]. Available: http://dx.plos.org/10.1371/journal.pone.0067346

[72] G. Erus *et al.*, "Imaging patterns of brain development and their relationship to cognition," *Cerebral Cortex*, vol. 25, no. 6, pp. 1676–1684, Jun. 2015. [Online]. Available: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgiDbfrom=pubmed&id=24421175&retmode=ref&cmd=prlinks, doi: 10.1093/cercor/bht425.

[73] J. Cheng *et al.*, "Brain age estimation from MRI using cascade networks with ranking loss," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3400–3412, Dec. 2021.

[74] 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. 2021.

[75] 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.