Brain Age Estimation based on sMRI via Adaptive Ensemble Learning

Zhaonian Zhang, Richard Jiang, Bryan Williams, Ce Zhang, Chang-tsun Li and Azeddine Beghdadi

***Abstract*—Predicting brain age is needed urgently in the biomedical domain. People began to realize that contrasting physical (real) age and predicted brain age can help to highlight brain issues and judge if patients' brains are healthy or not. Such age prediction is challenging for single model-based prediction. 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 (SVM), 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 this age-adaptive ensemble method, the weights of each model are learned automatically as a function of the age factor, 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) between the predicted age and the actual age, with the least MAE representing the highest accuracy in age prediction. By testing on the Predictive Analysis Challenge 2019, our novel ensemble model has achieved a MAE up to 3.19, which is a significantly increased accuracy in this brain age competition. If deployed in the real world, it could help doctors detect brain diseases more accurately and quickly, help pharmaceutical companies develop brain disease drugs faster, and make obvious progress in the field of brain science.**

***Index Terms*—Brain Age, Biomarks Ensemble Learning, Deep Learning.**

I. Introduction

The increasing aging population presents many important 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 [1-3]. 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 [4]. Although brain aging is a natural process, there are individual differences in the changes of brain volume, cortical thickness, and white matter microstructure [5-7]. 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 [8-9]. 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 [4].

The accurate prediction of brain age has not only critical scientific significance but also extensive clinical value [10]. 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 [11]. Besides, types of neurological diseases and metabolic diseases are closely related to abnormal aging of the brain [12], brain age prediction can effectively assist in providing clinical detection and condition tracking methods for these diseases.

Predicting brain age can also play a meaningful role in medical development, with clinical trials being an important part of clinical science [13-14]. At present, many pharmaceutical companies across the world are committed to the research of medicine for the treatment of age-related diseases, but 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 [14]. 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 [14]. 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 [4]. Existing research has shown that the difference between predicted brain age and the participant's

This paper was submitted on xx/xx/2021 for review, and accepted on xx/xx/2021. This work was supported in part by the EPSRC grant (xxxxxxx) and Leverhulme Trust Grant (xxxxxxxx). *(Correspondence Author: Richard Jiang, e-mail: r.jiang2@lancaster.ac.uk)*

Z. Zhang, R. Jiang and B. Williams are with School of Computing and Communications InfoLab21. Lancaster University Lancaster United Kingdom LA1 4WA.

R. Jiang and C. Zhang are with Lancaster Intelligent, Robotic and Autonomous Systems Centre (LIRA), Lancaster University, Lancaster, LA1 4YW, United Kingdom.

C. Zhang is with Centre of Excellence in Environmental Data Science (CEEDS), Lancaster Environment Centre, Lancaster LA1 4YQ, United Kingdom.

Chang-Tsun Li is with Deakin University, 75 Pigdons Rd, Waurn Ponds VIC 3216, Australia.

actual age is small for healthy people [4],[8],[11]. 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 [4], we can get that model by training deep learning models on healthy samples. And then, we would compare the predicted age and real age, if a sample's brain-predicted age is older than his real age, it represents poor brain health. Note, the training data must from healthy people, because the age predicted by the model shows what they should have for the healthy person under this brain condition. If the training data contains samples with diseases, the predicted age will not represent the age that the patient should have in this brain condition, so the comparison between the predicted age and the true age is meaningless.

Brain age estimation is kind of typical image classification task [1],[15]. There are many machine learning methods for it. Previously, Huang et al. [43] have applied CNNs in brain age estimation, and notably, Cole et al. [44] implemented a 3D CNN, which is trained on T1-weighted MRI, to predict brain age and achieved promising results. Figure 1 shows the process of brain age estimation. Our initial motivation is to identify which of the various models for predicting brain age can work 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 is that we proposed a novel non-linear 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 brain of each age group thus achieving high accuracy in predictions. Here, we include four different independent models, including a GoogLeNet, a ResNet, an SVM, and a self-designed CNN model. Non-linear age-adaptive developed in our ensemble model, utilize the changed weight of the independent model based on the age of the sample, where the model is adaptable to age changes and learns the brain characteristics over different ages.

We tested on the PAC 2019 competition dataset, and benchmarked our nl-AAE model amongst independent models. Such developed 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.

II. Related Work

In the traditional way, when doctors want to perform brain age prediction, they have to extract features from brain MRIs followed by classification or regression analysis, where information might be lost because the manually engineered features are likely not designed explicitly to include the most relevant information to brain age. When pre-processing the image, people always make numerous additional assumptions for any pre-processing pipelines, which are usually not met and such pre-processing can increase sources of error or uncertainties [18-19]. Extraction of features is time-consuming, while the doctor needs to make a decision within a few minutes to avoid the delay of treatment. Such an issue used to be an important reason why brain age prediction was not widely adopted.

Recently, the emergence of deep learning methods (such as CNNs) has attracted much interest widely across research and industrial communities. These methods are particularly powerful for image processing and computer vision, and have shown great potential for brain age prediction, which should increase the efficiency for doctors' consultation, clinical diagnosis as well as treatment decision making [3],[20-21]. At present, deep learning not only has successfully developed in the field of diagnosing schizophrenia [22], ADHD [23], autism [24], and Alzheimer's disease [25], but also helps to identify new biomarkers [26] and formulating new hypotheses [27].

Although deep learning has achieved success in biomedical fields, it still faces several challenges, both in terms of technology and practical applications [4],[28-30]. For example, deep neural networks require large sample sizes for fitting models, while neuroimaging datasets often have relatively small sample sizes [31-32]. The data scarcity has restricted the ability to learn image features effectively, and the problem of overfitting can also appear. Compared with 2D neuroimaging

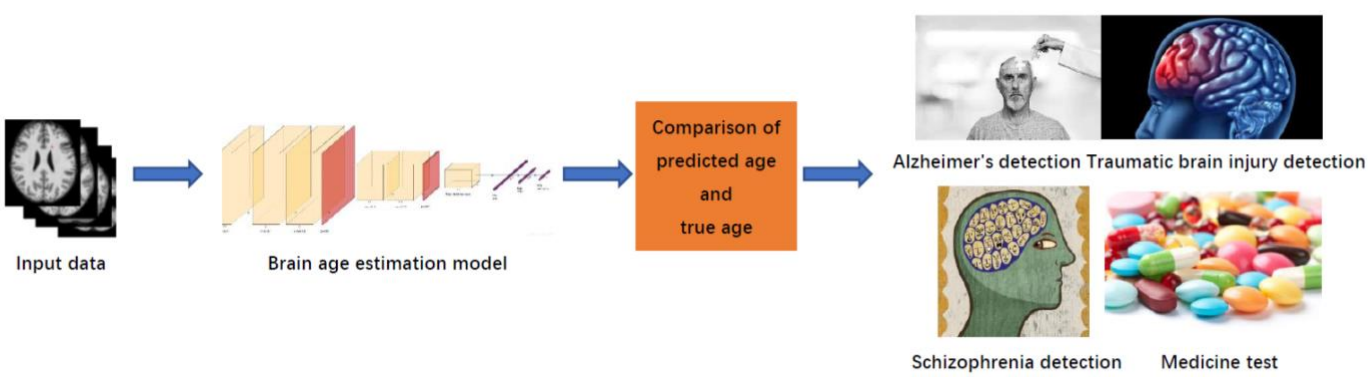

Figure 1 The process of brain age estimation

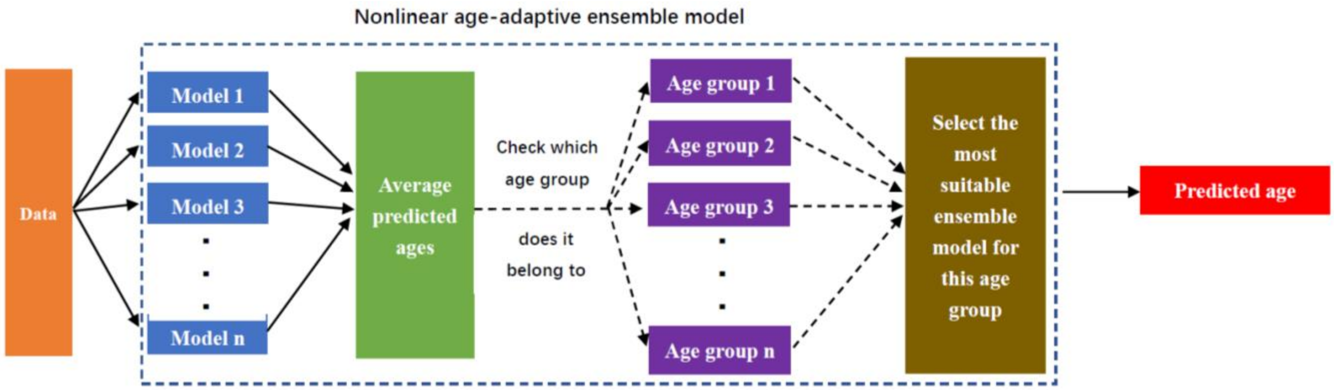

Figure 2 The estimation process of our ensemble model

data, 3D images require large GPU memory, which means that successful models in 2D data are not necessarily feasible in 3D scenes (e.g., ImageNet classification [33-34]). At present, there are challenges to further improve the prediction accuracy of deep learning models, and some studies have shown that in some neuroimaging data sets, even simple models could perform better than deep learning models [35-36]. Until now, how to choose the suitable complexity of the model is still an open question in scientific community.

Ensemble modeling is an alternative to choosing the best predictive model in machine learning. An ensemble model consists of a combination of predictions from several models to make the final prediction, by which the performance of model is increased [37]. There are several ways to combine individual models as an ensemble model, such as averaging, voting, to improve performance. As early as 1785, Marquis de Condorcet argued that if the probability of each independent voter being correct is above 0.5, then the addition of more voters increases the probability of the majority vote being correct [38]. This is good evidence to show that ensemble models have better performance than individual models.

III. Preliminaries on Brain Age Estimation

A. *Basic Independent Models*

Before we propose ensemble model, we need basic models as our blocks. The basic independent models for age estimation are detailed below. The reason why we choose them here is their performance for brain age estimation is good in previous researches [4,43-44].

1) *Convolutional Neural Networks*: The CNN we built here was implemented using Keras with TensorFlow as backend. For the first 5 consecutive blocks, each of them consists of a 3×3×3 3D convolution layer, a Batch Norm layer, a Max Pooling layer and an ELU activation. As for the 6th block, it contains one 1×1×1 3D convolution layer, a Batch Norm layer and an ELU activation. The 7th block contains an average pooling layer, a dropout layer, a 1×1×1 3D convolution layer and a softmax layer. The input data is a 3D volume image of 121×145×121 pixels, and the convolutional part of this model reduces this image to 128 feature maps of size 4×5×4. The finally fully connected layer reduces the feature maps down to the numbers that stand for predicted ages.
2) *GoogLeNet (Inception V1)*: GoogLeNet has many versions. Here we employed Inception V1 for the brain age estimation. The architecture of Inception V1 follows the literature but the convolution filters are altered to 3D. We kept the backbone the same but train the last 3 layers for the transfer learning for age estimation.
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 3×3×3 and stride 2×2×2, 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 1×1×1 and kernel size 3×3×3, 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 128×4×5×4=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.
4) *SVM*: SVM is a classical machine learning model, and its basic model is a linear classifier with the largest interval defined in the feature space. In this paper, we combined multiple two-class classification problems to achieve multi-classification for our brain age estimation.

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 can also be called multi-classifier system or committee-based learning. The general structure of ensemble learning 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.

There are six common types of ensemble learning, they are 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. 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. It will first train 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 training samples that the previous base learner did wrong will receive more attention in the follow-up, and then train the next base learner based on the adjusted sample distribution. This process is repeated until the number of base learners reaches the pre-specified value, and these base learners are finally combined with weight.

Bagging is the most famous representative of parallel type ensemble learning. Its principle is based on bootstrap sampling. Given a data set containing m samples, firstly, a sample will be randomly taken out and be put into the sampling set, and then the sample will be put back into the initial data set, so that this sample may still be selected in the next sampling, after m random sampling operation, a sampling set containing m samples will be finished. Assuming that the number of base learners required in ensemble learning is T, we can generate T sample sets, each containing m training samples, then we can train a base learner based on each sample set, and finally combine these learners. Random Forest is one of the most famous extended variants of Bagging.

BMA, BMC and Stacking represent different model combining strategies. BMA uses the weighted average method to combine the models, and the weight of the model is equal to the posterior probability of the model. BMC 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. Stacking is usually better than BMA and BMC because it is more robust, and BMA and BMC are sensitive to model approximation errors.

B. *Nonlinear Age-Adaptive Ensemble Model*

Through experimentations, we found that the performance of all the models is influenced by the sample's true age (see part V). This shows that some models are suitable for predicting young samples, and some models are suitable for older samples. In order to make the prediction results better, we built a model called nonlinear age-adaptive ensemble model, which is similar to Stacking, and the difference is that our ensemble model can adjust weights of inside independent models according to age.

First, we used 4 different independent models as its initial learners, they are ResNet, GoogLeNet, CNN and SVM. 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. Besides,

$$\sum_i \omega_i = 1$$

Selecting the weights of independent models is important, now we will show the methods we used here. In each age group, we

Begin

Input *brain sMRI* **as** x

$H_1 = ResNet(x)$

$H_2 = GoogLeNet(x)$

$H_3 = CNN(x)$

$H_4 = SVM(x)$

$H_{ave} = \frac{1}{4} * (H_1 + H_2 + H_3 + H_4)$

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

Figure 3 Pseudo list of brain age estimation

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

Next, we need to minimize the value of the loss function, here, we used the gradient descent method to complete this task, its equation is shown below:

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

$$\omega = \omega - \lambda H^T(H\omega - Y). \quad (3)$$

(∂ (not convenient symbol / very often as partial derivative operator …i suggest to replace by λ)) λ is the learning rate. After several iterations (We should specify the number of iteration; Moreover, it is implicitly assumed that the process converges. Is this really the case? Can we demonstrate it? it is not clear)), we got the final result of ω .

The least squares method can also accomplish this task. It can be described as below: The solution is given by:

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

Note, the results from these two methods are the same after experiments. UNCLEAR.. Rephrase .. Explain

In this way, for each age group, we got a series of suitable weights (What is meant by suitable ? in terms of what? According to which criteria?) for independent models, so that we have an ensemble model for each age group, and these ensemble models constitute the final nonlinear age-adaptive ensemble model, which can be expressed as:

$$F(x) = \sum\nolimits_{A \in age} \omega_A H(x, p_A) \quad (5)$$

age represents a collection 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 $(H_1, H_2, ..., H_n)$, after that we calculate the average of all predicted ages and call it H_{ave}, $H_{ave} = \frac{1}{n}(H_1 + H_2 + ... + H_n)$. (What is the reason and the idea behind taking such a simple arithmetic average. There is no reason to give the same importance to the H's) Next, we check which age group does the H_{ave} belong to, we have several age groups $(G_1, G_2, ..., G_n)$ and there is an ensemble model in each age group $(M_1, M_2, ..., M_n)$, 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 2 and Figure 3 show the estimation process of our ensemble model.

V. Experimental Results

A. Experimental Results from the Proposed AAE Model

The test method we used here is 5 times 5-fold-cross-validation, with the mean MAE of them as final results. In our first experiment, we divided the prediction results of four independent models into two parts, one contains the samples over 40 years old, and the other contains the samples under 40 years old. We separately trained models on two sets of data, used the 4 models' results as secondary learner's sample input feature, established two ensemble models, and combined them into a non-linear ensemble model, and its final average MAE is 3.45.

On the second experimental, 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. This time, the ensemble model's average MAE is 3.39.

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. As the age increases, the amount of data becomes less and less, therefore, for the 30 to 60-year-old samples, we took every 5 years old samples as a group, then we used the 60 to 70-year-old samples as a group, and the 70 to 90-year-old samples as a group. Through this division, the ensemble model's performance is the best whose average MAE is 3.19. The results of all models are shown in Table I.

B. Investigation on performance per Individual Models

In this part, we will investigate the performance of each individual model. As we mentioned before, in this project, we used MAE as the standard for evaluating the model. The smaller the MAE, the better the model would be. And for each model, we tested it by 5 times 5-fold cross-validation, and calculated their mean values as this model's final performance.

Figure 4 suggests Ensemble Models as the best of all models and the performance of 'mean' and 'median' is similar. The performance of SVM is not good as expected since SVM is a traditional shallow machine learning approach compared to CNN. GoogLeNet always performs best, followed by ResNet, but both models perform better than 6-layer CNN.

C. Investigation on Age-Sensitivity per Models

The age-sensitivity shows the trend of MAE for each model according to age. Therefore, in this part, we investigate the age-sensitivity of the models we built before. Figure 5 shows the results, it says that all the models are good at predicting young people but not old people, the MAE 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 SVM 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 the independent models', but it is 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 the reasons of machine learning side, this is because when we are training the model, the number of older samples is smaller and the training is insufficient, which causes the model to predict the older samples inaccurately. For the reasons of medical side, we think this is because as age grows, the differences between the brains of different people will become larger, which will increase the difficulty of prediction, and the brain differences between young people are not that big. So that when training the model in the future, we can increase the proportion of young people's data to improve the accuracy of the model.

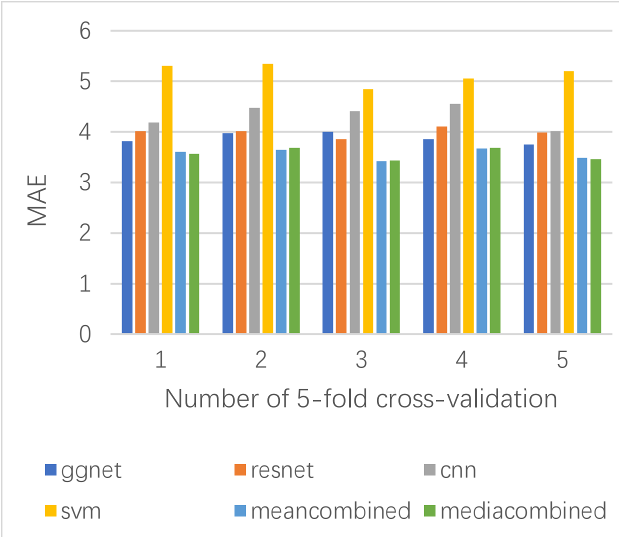

Figure 4 Performance of each individual model

ggnet: GoogLeNet resnet: ResNet
cnn: the self-defined CNN SVM: SVM
meancombined: an Ensemble Model which takes the average of the 4 model predictions as the prediction result
mediancombined: an Ensemble Model which takes the median of the 4 model predictions as the prediction result

Model	Average MAE
ResNet	3.99
SVM	5.15
CNN (trained by all the data)	4.33
GoogLeNet	3.88
Mean-combined Ensemble Model	3.57
Median-combined Ensemble Model	3.72
4 models combined Ensemble Model	3.52
4 models combined AAE (divide age into 2 parts)	3.45
4 models combined AAE (divide age into 6 parts)	3.39
4 models combined AAE (divide each age as a group)	3.19

Table I Results of all the models

D. Learning the Model Weights

Figure 6 shows the change of each independent model's weights in the AAE according to age. Through this figure, we can know the different importance of each independent model in the AAE at different ages.

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

GoogLeNet model is most important in the AAE. When the sample age is 20 to 40 years old, the effect is small, but in other age groups, especially for middle-aged and elderly people over 50 years old, the weight is large. This shows that the function of GoogLeNet is powerful, and it is suitable to predict the age of middle-aged and elderly samples.

As for ResNet, it 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, its performance is average, so it does not hold much 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.

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

VI. Discussion

A. Discussion of AI models

In this paper, the most important contribution is the nonlinear age-adaptive ensemble (AAE) model *first* proposed in BrainAge, and showed better results than other benchmark models. 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, which makes AAE become a dynamic model, where the prediction results are more accurate given BrainAge projection is sensitive to age change. However, we still found defects and improvements in the process of the experiment. For the choice of initial models, we could try other advanced methods. For example, as for GoogLeNet, we used the Inception V1 version, now it has already had Inception V4 versions.

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 should be combined with various factors.

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 the development of this field is to build a brain age model based on large image sets that completes large data training. And can apply the model to practice to provide accurate personalized cloud diagnosis services.

B. Discussion of medical findings

According to our research, we also have some interesting findings from medical perspectives. First, we found that the performance of models will decrease with the age of the sample, which shows that the brains of young people are similar, and along with the increase of age, the risk of people suffering from brain diseases increases, the differences between the brains are getting larger, and the accuracy of the model's prediction begins to decline.

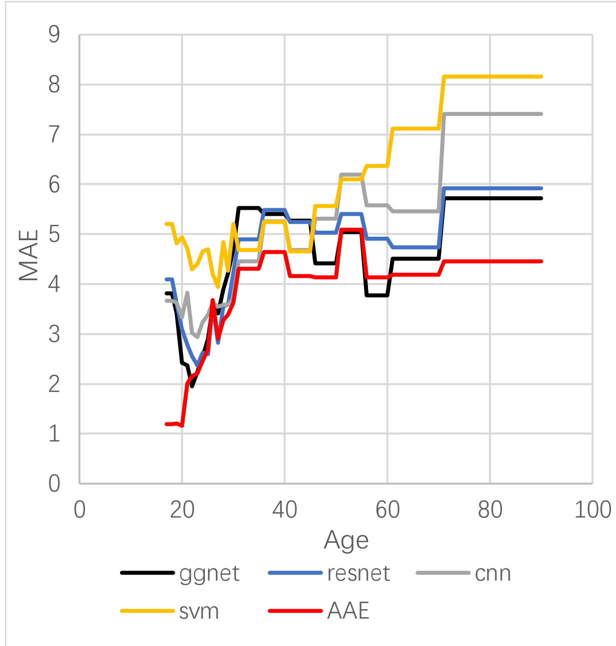

Figure 5 Age-Sensitive of models

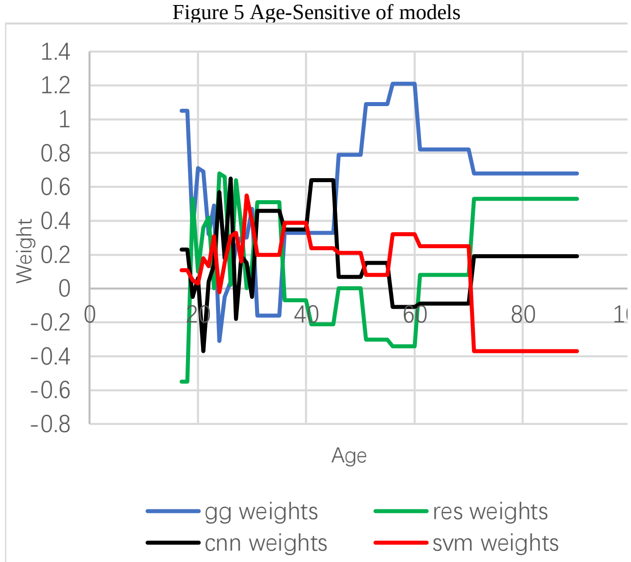

Figure 6 Individual models' weights in AAE

From the results of our experiments, we believe that changes in the brain can be divided into 4 stages, they are 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 can see 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 London, researchers explained that our brains slowly transition to adulthood, which is finally reached in our 30s. And after the age of 30, the brain's working memory capacity begins to slowly decline [41], this is the same as our research results. 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 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 [42]. At the age of 50-70, the brain does not change much, but after the age of 70, the brain will have the last big 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 [41], 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 [8],[9]. 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 by the University of Miami in Florida [39] 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.

In 2017, researchers from Albert Einstein College of Medicine in New York City, NY, found that stem cells in the brain's hypothalamus likely control how fast aging occurs in the body [40]. 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 Model, which is combined by 4 models and achieved state-of-the-art brain age prediction using T1-weighted structural MRI images. Its MAE is 3.19, which is a great result for brain age prediction. As for independent models, we demonstrated that GoogLeNet and ResNet are more suitable for BrainAge project, and SVM's performance is the worst. Besides, we also found suitable age groups for each model, the results show GoogLeNet basically has a good performance on data of all age groups, especially for middle-aged and old samples, ResNet and CNN are suitable for predicting the age of young samples but the performance on middle-aged and old samples is not good, while the performance of SVM is generally average, but it is not suitable for predicting the age of young samples. Further, the experimental results show that the brain basically has 4 stages according to age, the older the person, the greater the brain age gap, and the greater the prediction error.

We hope that the methods we have used here can inspire some ideas and provide references for other researchers to study BrainAge further.

References

[1] Reeve, A., Simcox, E. & Turnbull, D. Ageing and Parkinson's disease: why is advancing age the biggest risk factor? Ageing Res. Rev. 14, 19–30 (2014)

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

[3] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

[4] Cole J H, Franke K. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends Neurosci, 2017, 40(12): 681-690

[5] Groves, A.R., Smith, S.M., Fjell, A.M., Tamnes, C.K., Walhovd, K.B., Douaud, G., Woolrich, M.W., Westlye, L.T., 2012. 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 63, 365–380. http://dx.doi.org/10.1016/j.neuroimage.2012.06.038, (URL⟨http://www.sciencedirect.com/science/article/pii/S105381191200 6532⟩)

[6] Storsve, A.B., Fjell, A.M., Tamnes, C.K., Westlye, L.T., Overbye, K., Aasland, H.W., Walhovd, K.B., 2014. Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: regions of accelerating and decelerating change. J. Neurosci. 34,8488–8498,(URL⟨http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.039 1-14.2014⟩)

[7] Fjell, A.M., Walhovd, K.B., Fennema-Notestine, C., McEvoy, L.K., Hagler, D.J., Holland, D., Brewer, J.B., Dale, A.M., 2009. One-year brain atrophy evident in healthy aging. J. Neurosci. 29, 15223–15231. http://dx.doi.org/10.1523/JNEUROSCI.3252-09.2009

[8] Luders, E., Cherbuin, N., Gaser, C., 2016. Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. NeuroImage 134, 508–513. http://dx.doi.org/10.1016/j.neuroimage.2016.04.007

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

[10] Castellanos, F.X., Di Martino, A., Craddock, R.C., Mehta, A.D., Milham, M.P., 2013. Clinical applications of the functional connectome. NeuroImage 80,527–540. http://dx.doi.org/10.1016/j.neuroimage.2013.04.083,(URL⟨http://eutils.n cbi.nlm. nih.gov/entrez/eutils/elink.fcgi dbfrom=pubmed & id=23631991 & retmode=ref & cmd=prlinks⟩)

[11] Cole, J.H., Ritchie, S.J., Bastin, M.E., Valdés Hernández, M.C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., Wray, N.R., Redmond, P., Marioni, R.E., Starr, J.M., Cox, S.R., Wardlaw, J.M., Sharp, D.J., Deary, I.J., 2018. Brain age predicts mortality. Mol. Psychiatry 23, 1385–1392. https://doi.org/10.1038/mp.2017.62

[12] The Brainstorm Consortium, Anttila, V., Bulik-Sullivan, B., Finucane, H. K. & Walter, R. K. Analysis of shared heritability in common disorders of the brain. Science 360, 6395 (2018)

[13] Bzdok, D., Varoquaux, G., Steyerberg, E.W., 2020. Prediction, not association, paves the road to 322 precision medicine. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2020.2549

[14] Jenni A. M. Sidey-Gibbons and Chris J. Sidey-Gibbons,2019.Machine learning in medicine: a practical introduction. Sidey-Gibbons and Sidey-Gibbons BMC Medical Research Methodology (2019) 19:64 https://doi.org/10.1186/s12874-019-0681-4

[15] Bron, E.E., Smits, M., van der Flier, W.M., Vrenken, H., Barkhof, F., Scheltens, P., Papma, J.M., Steketee, R.M.E., Méndez Orellana, C., Meijboom, R., Pinto, M., Meireles, J.R., Garrett, C., Bastos-Leite, A.J., Abdulkadir, A., Ronneberger, O., Amoroso, N., Bellotti, R., Cárdenas-Peña, D., Álvarez-Meza, A.M., Dolph, C.V., Iftekharuddin, K.M., Eskildsen, S.F., Coupé, P., Fonov, V.S., Franke, K., Gaser, C., Ledig, C., Guerrero, R., Tong, T., Gray, K.R., Moradi, E., Tohka, J., Routier, A., Durrleman, S., Sarica, A., Di Fatta, G., Sensi, F., Chincarini, A., Smith, G.M., Stoyanov, Z.V., Sørensen, L., Nielsen, M., Tangaro, S., Inglese, P., Wachinger, C., Reuter, M., vanSwieten, J.C., Niessen, W.J., Klein, S., Alzheimer's Disease Neuroimaging Initiative. 2015. Standardized evaluation of algorithms for computeraided diagnosis of dementia based on structural MRI: the CADDementia challengeNeuroImage 111, 562–579. http://dx.doi.org/10.1016/ j.neuroimage.2015.01.048, (URL ⟨http://linkinghub.elsevier.com/retrieve/pii/ S1053811915000737⟩)

[16] Reeve, A., Simcox, E. & Turnbull, D. Ageing and Parkinson's disease: why is advancing age the biggest risk factor? Ageing Res. Rev. 14, 19–30 (2014)

[17] Abbott, A., 2011. Dementia: A problem for our age. Nature 475, S2-S4

[18] Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating

automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12, 26–41. http://dx.doi.org/ 10.1016/j.media.2007.06.004

[19] Liu, X., Niethammer, M., Kwitt, R., Singh, N., McCormick, M., Aylward, S., 2015. Lowrank atlas image analyses in the presence of pathologies. IEEE Trans. Med. Imaging 34, 2583–2591. http://dx.doi.org/10.1109/TMI.2015.2448556

[20] De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, C.O., Raine, R., Hughes, J., Sim, D.A., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, P.T., Suleyman, M., Cornebise, J., Keane, P.A., Ronneberger, O., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat. Med. 24, 1342–1350. https://doi.org/10.1038/s41591-018-0107 6

[21] Kohl, S.A.A., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K.H., Ali Eslami, S.M., Rezende, D.J., Ronneberger, O., 2018. A probabilistic U-net for segmentation of ambiguous images, in: Advances in Neural Information Processing Systems. pp. 6965–6975

[22] Zeng, L.L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., Chen, X., Liu, Z., Yin, H., Tan, Q., Wang, K., Hu, D., 2018. Multi-Site Diagnostic Classification of Schizophrenia Using Discriminant Deep Learning with Functional Connectivity MRI. EBioMedicine 30, 74–85. https://doi.org/10.1016/j.ebiom.2018.03.017

[23] Zou, L., Zheng, J., Miao, C., McKeown, M.J., Wang, Z.J., 2017. 3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI. IEEE Access 5, 23626–23636. https://doi.org/10.1109/ACCESS.2017.2762703

[24] Thomas, R.M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., van Wingen, G., 2020. Classifying Autism Spectrum Disorder Using the Temporal Statistics of Resting-State Functional MRI Data With 3D Convolutional Neural Networks. Front. Psychiatry 11, 1. https://doi.org/10.3389/fpsyt.2020.00440

[25] Liu, M., Zhang, J., Adeli, E., Shen, D., 2018. Landmark-based deep multi-instance learning for brain disease diagnosis. Med. Image Anal. 43, 157–168. https://doi.org/10.1016/j.media.2017.10.005

[26] Rosenberg, M.D., Casey, B.J., Holmes, A.J., 2018. Prediction complements explanation in understanding the developing brain. Nat. Commun. 9, 1–13. https://doi.org/10.1038/s41467-018-02887-9

[27] Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25, 289–310. https://doi.org/10.1214/10- STS330

[28] Kawahara, J., Brown, C.J., Miller, S.P., Booth, B.G., Chau, V., Grunau, R.E., Zwicker, J.G., Hamarneh, G., 2017. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. Neuroimage 146, 1038–1049. https://doi.org/10.1016/j.neuroimage.2016.09.046

[29] Arslan, S., Ktena, S.I., Glocker, B., Rueckert, D., 2018. 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. pp. 3–13. https://doi.org/https://doi.org/10.1007/978-3 030-00689-1_1

[30] Baumgartner, C.F., Koch, L.M., Tezcan, K.C., Ang, J.X., 2018. Visual Feature Attribution Using Wasserstein GANs, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp. 8309–8319. https://doi.org/10.1109/CVPR.2018.00867

[31] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y

[32] Raghu, M., Zhang, C., Kleinberg, J., Bengio, S., 2019. Transfusion: Understanding Transfer Learning for Medical Imaging, in: Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 3347–3357

[33] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems. pp. 1097–1105

[34] Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv Prepr. arXiv1409.1556

[35] He, T., Kong, R., Holmes, A.J., Nguyen, M., Sabuncu, M.R., Eickhoff, S.B., Bzdok, D., Feng, J., Yeo, B.T.T., 2019. Deep Neural Networks and Kernel Regression Achieve Comparable Accuracies for Functional Connectivity Prediction of Behavior and Demographics. bioRxiv 473603. https://doi.org/10.1101/473603

[36] Schulz, M., Yeo, B.T.T., Vogelstein, J.T., Mourao-, J., Kather, J.N., Kording, K., Richards, B., Bzdok, D., 2019. Deep learning for brains?: Different linear and nonlinear scaling in UK Biobank brain images vs. machine-learning datasets. bioRxiv 757054

[37] Andreas Veit, Michael J Wilber, and Serge Belongie. Residual networks behave like ensembles of relatively shallow networks. In Advances in neural information processing systems, pages 550–558, 2016

[38] Marquis de Condorcet. Essay on the application of analysis to the probability of majority decisions. Paris: Imprimerie Royale, 1785

[39] Joshua Z. Willey, Hannah Gardener, Michelle R. Caunca, Yeseon Park Moon, Chuanhui Dong, Yuen K. Cheung, Ralph L. Sacco, Mitchell S.V. Elkind, Clinton B. Wright, 2016. Leisure-time physical activity associates with cognitive decline. The Northern Manhattan Study Neurology May 2016, 86 (20) 1897-1903; DOI: 10.1212/WNL.0000000000002582

[40] Zhang, Y., Kim, M., Jia, B. et al. Hypothalamic stem cells control ageing speed partly through exosomal miRNAs. Nature 548, 52–57 (2017). https://doi.org/10.1038/nature23282

[41] Guo JY, Isohanni M, Miettunen J, Jääskeläinen E, Kiviniemi V, Nikkinen J, Remes J, Huhtaniska S, Veijola J, Jones PB, Murray GK (2016), "Brain structural changes in women and men during midlife" *Neuroscience Letters* 615, pp. 107-112

[42] Francine Grodstein. How early can cognitive decline be detected?.BMJ 2012;344:d7652

[43] Huang, T.-W. et al. Age estimation from brain MRI images using deep learning. In Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on Biomedical Imaging, 849–852 (IEEE, 2017)

[44] Cole, J. H. et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. NeuroImage 163, 115–124 (2017)

[45] C. Chiang, C. Barnes, P. Angelov and R. Jiang, "Deep Learning based Automated Forest Health Diagnosis from Aerial Images", 2020, IEEE Access.

[46] R. Jiang and D. Crookes, "Shallow Unorganized Neural Networks Using Smart Neuron Model for Visual Perception", 2019, IEEE Access.

[47] G Storey, R Jiang, A Bouridane, C T Li, "3DPalsyNet: A Facial Palsy Grading and Motion Recognition Framework using Fully 3D Convolutional Neural Networks", 2019, IEEE Access.

[48] Jiang, Z., Chazot, P. L., Celebi, M. E., Crookes, D. & Jiang, R., "Social Behavioral Phenotyping of Drosophila with a 2D-3D Hybrid CNN Framework", 2019, IEEE Access.

[49] R Jiang, ATS Ho, I Cheheb, N Al-Maadeed, S Al-Maadeed, A Bouridane, "Emotion recognition from scrambled facial images via many graph embedding", 2017, Pattern Recognition.

[50] R Jiang, S Al-Maadeed, A Bouridane, D Crookes, ME Celebi, "Face recognition in the scrambled domain via salience-aware ensembles of many kernels", 2016, IEEE Transactions on Information Forensics and Security.

[51] R Jiang, D Crookes, N Luo, MW Davidson, "Live-cell tracking using SIFT features in DIC microscopic videos", 2010, IEEE Transactions on Biomedical Engineering.