Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
December 2021, vol. 26, no. 2, pp. 116–121
https://doi.org/10.2478/acss-2021-0014
https://content.sciendo.com

Curriculum Learning for Age Estimation from Brain MRI

Alican Asan[1], Ramazan Terzi[2], Nuh Azginoglu[3*]
[1]*Department of Big Data and Artificial Intelligence, Digital Transformation Office, The Presidency of the Republic of Turkey, Ankara, Turkey*
[2]*Department of Computer Engineering, Amasya University, Amasya, Turkey*
[3]*Department of Computer Engineering, Kayseri University, Kayseri, Turkey*

***Abstract* – Age estimation from brain MRI has proved to be considerably helpful in early diagnosis of diseases such as Alzheimer's and Parkinson's. In this study, curriculum learning effect on age estimation models was measured using a brain MRI dataset consisting of normal and anomaly data. Three different strategies were selected and compared using 3D Convolutional Neural Networks as the Deep Learning architecture. The strategies were as follows: (1) model training performed only on normal data, (2) model training performed on the entire dataset, (3) model training performed on normal data first and then further training on the entire dataset as per curriculum learning. The results showed that curriculum learning improved results by 20 % compared to traditional training strategies. These results suggested that in age estimation tasks datasets consisting of anomaly data could also be utilized to improve performance.**

***Keywords* – Age estimation, age prediction, brain MRI, curriculum learning.**

I. Introduction

Estimating the age of individuals from features such as human facial details, teeth, DNA, speech signals and brain structure is one of the fields of study that has applications [1]–[3]. Related studies are justified for forensic or disease diagnosis and are usually performed using image analysis techniques [2], [4].

Advances in artificial intelligence studies and convolution-based architectures have led to improved machine learning and deep learning-based studies on age prediction from MRI images. In this area, some studies performed age analysis from MRI images of bones such as clavicle [5], [6], radius, ulna [4], femur [7] and wrist [8], as well as MRIs of the hand [5], [9] and knee [10]. In addition, age estimation can be made from the data including teeth [5], [11] and dentine images [12]–[14].

One of the most important methods used for age estimation is the detection of morphological changes in different brain regions from brain MRI images. Thus, age estimation can be made through age-related changes seen in anatomical structures in the brain [15]–[17].

Lam et al. [18] estimated age using T1-weighted brain MR images to predict Alzheimer's, and Behesthi et al. [19] also performed a similar study for diagnosing both Alzheimer's and Parkinson's. While Lam et al. [18] used 3D-CNN as a method identical to that of Ueda et al. [20]. Behesthi et al. used 3D Convolutional Neural Network (3D-CNN) with the Support Vector Regression algorithm [19]. Huang et al. sought a solution to the same problem with the VGG method [21]. Afshar et al. preferred Extreme Learning Machine as the method of choice [1]. In [22], an Age-Net framework is developed, which first performs chronological age estimation with the Inception v1 CNN architecture and then estimates biological age with SqueezeNet [22]. Su et al. made a sparse representation for age prediction and age-related classification [17]. While Franke et al. [23] used the Relevance Vector Machine for regression, Wang et al. [3] used Hidden Markov Models. Apart from the above-mentioned research, there are also studies discussing the use of transfer learning techniques for age estimation [24], [25].

Curriculum learning in age estimation has also been discussed [26]–[28]. In the curriculum learning approach, the system does not learn from random data, unlike the classical education model. Instead, the system is trained with easy examples first, followed by more difficult ones. This approach has been developed by taking advantage of the observation that living beings learn better and faster when they start with easy patterns, and it is preferred because of its ability to generalize [26]. Furthermore, we come across multiple studies devoted to curriculum learning-based age estimation. Wang et al. [27] estimated age from face images by using the ResNet-34 architecture; Li et al. made age estimates from face pictures, using a cumulative hidden layer learning model as a deep learning architecture [28]. Moreover, Kim et al. predicted the age using dental X-ray images through DenseNet-121 architecture [29]. As can be seen, although there are studies that use different data types in various fields in which curriculum learning is used in age estimation, it has not been used in age estimation from brain MRI.

In this study, three different learning strategies for age prediction from brain MRIs were experimentally tested on

* Corresponding author's e-mail: nuhazginoglu@kayseri.edu.tr

©2021 Alican Asan, Ramazan Terzi, Nuh Azginoglu.
This is an open access article licensed under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), in the manner agreed with Sciendo.

3D CNN architecture to measure the impact of curriculum learning on estimating the age from brain MRI images. In the first strategy, the model was trained with healthy brain MRIs, while brain MRIs containing anomalies were used together with the normal data in the second strategy. In the last strategy, the system was trained first with healthy brain MRIs and then with both healthy and anomaly MRIs, and thus the curriculum learning method was applied to the model.

II. Material and Method

A. Dataset

In this research, Gazi Brains 2020 [30], which includes 50 patients diagnosed with high-grade glioma (HGG) and 50 individuals with no anomalies, was used to predict age from brain MRI. Only FLAIR weighted images were selected for training deep learning models. The first and last slices were removed, and the middle 17 slices were chosen for the research. The images were rescaled, resized and converted to grayscale. As predicting the age of the individual consists of the information from the entirety of the brain, 3D models were proposed. A 3D image of 224×224×17 was constructed from the slices to represent the entire brain.

The dataset has individuals with ages ranging between 7 and 81. The age histogram of the whole dataset is given in Fig. 1.

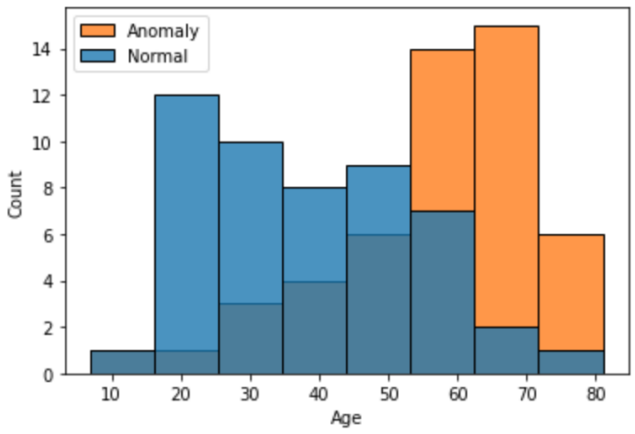


Fig. 1. Age distribution of Gazi Brains 2020 dataset.

The mean, median and standard deviation values for normal and anomaly data are given in Table I.

TABLE I
GAZI BRAINS 2020 DATASET STATISTICS

Dataset	Metrics		
	Mean	**Median**	**Standard Deviation**
Normal	38.68	38.00	15.65
Anomaly	57.34	60.50	15.18

As can be observed from the histogram, the anomaly dataset generally consists of elderly individuals. This difference between the age groups of normal and anomaly datasets should be taken into account when training and testing age detection models.

B. Learning Strategies

Three different learning strategies were used in the research. All learning strategies are presented in Fig. 2.

- **Strategy 1:** Classical Learning Strategy with only Normal Data. In this strategy, only normal data were used for training, validating and testing models.
- **Strategy 2:** Classical Learning Strategy with mixed data. In this strategy, both normal and patients with HGG data were used for training, validating. Models were trained from scratch with randomly shuffled data. Testing was performed with only normal data.
- **Strategy 3:** Curriculum Learning Strategy. In this strategy, firstly, models were trained using normal data and then training continued by adding anomaly data. Validating and testing were performed with only normal data.

C. 3D-CNN

When working with two-dimensional images, convolutions are applied to 2D feature maps [31]. In the brain MR images, a three-dimensional image formed by combining a certain number of slices emerges. Considering that there is a brain MRI and age information for every person in the dataset within the scope of our study, it is evident that a three-dimensional structure should be used. The age corresponding to the subject can be represented by all of the slices together rather than specific slices alone. For this reason, 3D CNN has been preferred as a deep learning architecture.

A custom 3D CNN model was built with a linear output for regression as shown in Fig. 3. The model consists of four convolutional blocks and two fully connected layers. Convolutional blocks consist of two convolutional layers with (2, 2, 2) kernel sizes followed by batch normalization and ReLu Activation Function followed by the Max Pooling layer. Max Pooling layers were set to (2, 2, 1) to conserve slice information, as it was not beneficial to perform max pooling on the slice dimension. Fully connected layers consist of 128 and 64 neurons, respectively, connected to a linear layer as output for regression (Fig. 3).

The data were structured for 3D using 2D slices and, after removing the first and last layers, 17 slices were stacked on top of each other to represent the entire brain. Therefore, the shape of the data was 224, 224, 17, 1. When performing max pooling on this kind of data structure, it was decided not to perform max pooling on the slice dimension and set the max pooling to (2, 2, 1). This way, the slice information is preserved without dimension reduction on this dimension while still performing max pooling on the image dimensions.

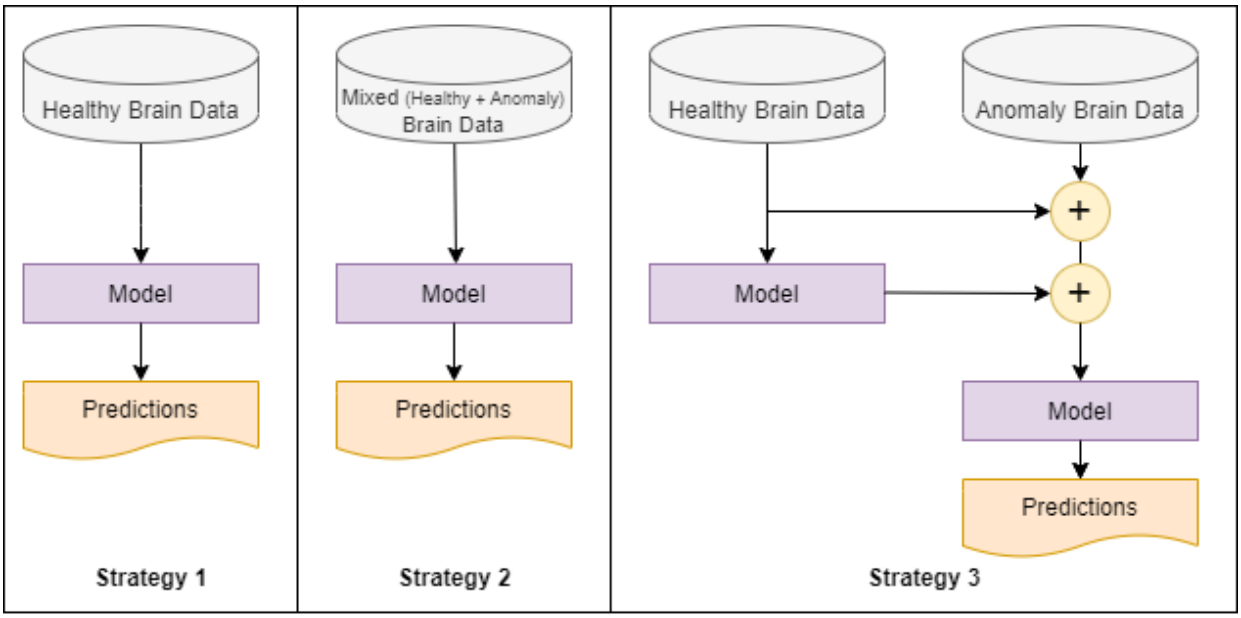


Fig. 2. The proposed learning strategies.

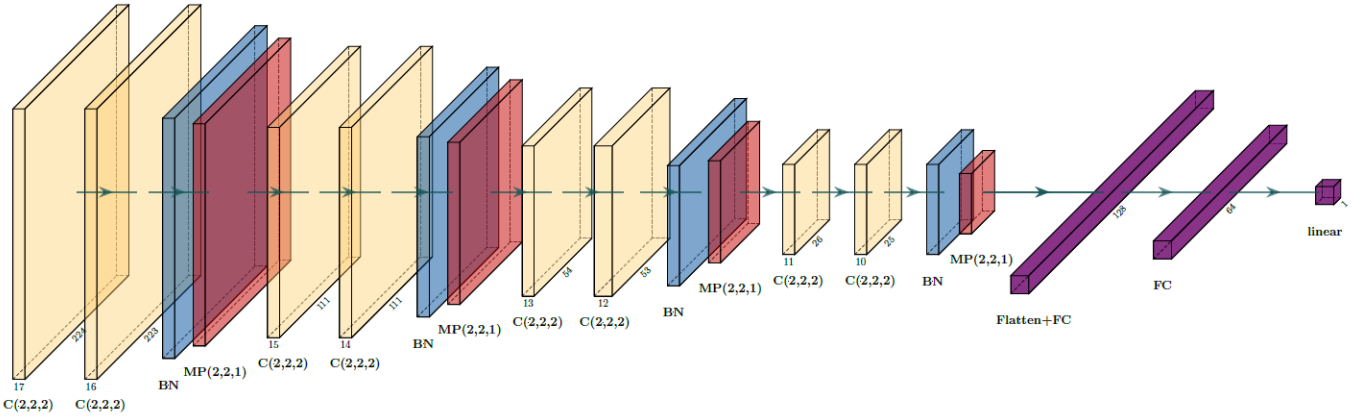


Fig. 3. 3D-CNN model architecture.

III. Results and Discussion

The experiments were performed with 5-fold cross-validation. For each fold, 3 parts were used for training, 1 part was used for validation and 1 part was used as a test set. The model was trained using the training set and the performance of the model was observed during the training according to the validation set. The model with the best validation performance was selected and its success on the test set was reported as experimental results.

A learning rate of 0.00008 was used with the Adam optimizer. A batch size of 4 with a dropout of 0.3 was selected and the models were trained for 100 epochs. The Mean Absolute Error (MAE) was used for both loss function and success metric. MAE value is one of the most frequently used metrics in regression problems because it is easily interpretable. It is a measure of the difference between two continuous variables. One of the variables used here is the actual value, and the other is the value predicted by the model. The MAE is the average vertical distance between each true value and the line that best fits the data and the average horizontal distance between each data point and the line that best fits the data. Thus, the MAE is a linear value that measures the mean size of errors in a set of estimates, with all individual errors weighted equally on the mean. An MAE value of zero is the ideal situation and indicates that all the predicted values agree with the true values. Also, all positive real numbers are probable MAE values, while the MAE value does not take values between specific numbers. When two or more models are compared, the lower MAE value is expressed as more successful. The calculation of the MAE value is given below (1).

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j| \ . \quad (1)$$

In Strategy 1, training validation and testing of the model are done only on normal data. Only half of the dataset consists of normal individuals MRI data, so half the data in this strategy goes unused.

In Strategy 2, training of the model is performed using both anomaly and normal data, and validating and testing are carried out on only normal data. Since the dataset consists of patients with anomalies and normal individuals, using the other half of the data improves results as the data have essentially doubled. However, this strategy might be unreliable as the patients with

anomalies tend to be older in the dataset and might cause a bias. Furthermore, the model can learn to detect age using the features from the tumors or anomalies. Curriculum learning strategy is proposed to overcome these challenges and to improve results on the dataset using as much data as possible.

In Strategy 3, using curriculum learning, the pre-trained model from the 1st step is taken and trained further with a dataset consisting of both normal data and anomaly data. This way it is made sure that the model has learned the right features to look for when it comes to age detection. Using curriculum learning made it possible for the entire dataset to be utilized, increasing the size of the data and helping against overfitting.

After 50 epochs of only being trained on the normal dataset, the best performing epoch model on the validation dataset of every fold was taken to be further trained on anomaly and normal data together. Starting from relatively easy data to find the right features and then performing training on a bigger dataset, it is possible to improve the results (see Table II).

Table II shows that the least successful model has been the one trained from only normal data; it is also the model that has only seen half of the dataset, therefore this model has the worst test results and is more prone to overfitting.

TABLE II
5-FOLD CROSS-VALIDATION MAE RESULTS OF LEARNING STRATEGIES

Folds	Classical learning with only normal data	Classical learning with mixed data	Curriculum learning
1	09.59	07.43	08.12
2	17.14	13.25	12.70
3	12.30	06.96	06.26
4	11.05	15.36	07.09
5	12.68	11.06	08.13
Avg.	12.55	10.81	08.49

Classical learning with mixed data strategy has improved the results on average of five folds by 14 %. This is mainly due to the size of the data increasing in the training set. It is entirely possible for this model to have more biases due to the nature of the anomaly dataset.

Curriculum learning strategy has improved the results on average of five folds by 20 %. This model is the least prone to overfitting and bias compared to the other two.

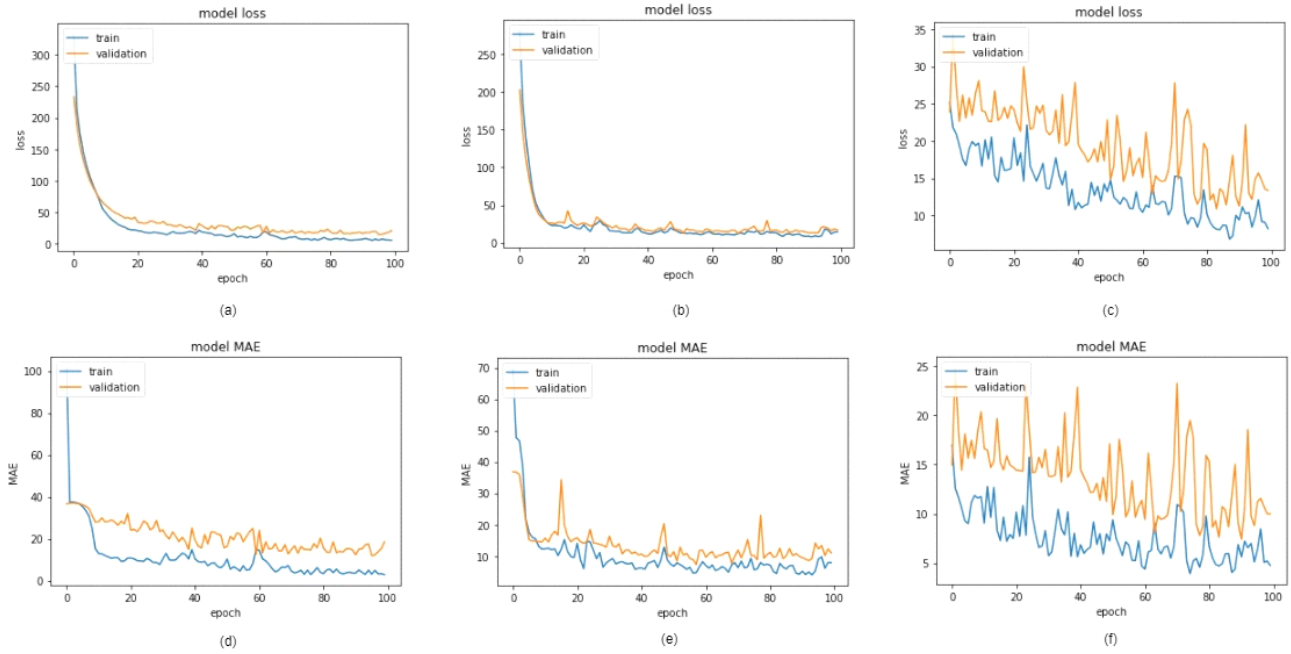


Fig. 4. Loss and MAE plots (a), (d) for Strategy 1 (b), (e) for Strategy 2 and (c), (f) for Strategy 3.

In Fig. 4, Y-axes on a, b and c graphs show loss values while d, e and f are MAE values, respectively, for Strategy 1, 2 and 3. The X-axis shows the number of epochs trained. The loss curves of the models have not overfit as the train and validation losses are roughly similar and while the models starting from scratch (Strategy 1 and 2) have very high loss values (200–300) and MAE values (60–80) from the start, the curriculum learning model (Strategy 3) starts from a lower loss (25–30) and MAE (15–20) as it has already been trained on normal data and therefore starts the task with better weights set already.

IV. CONCLUSION AND RECOMMENDATIONS

In this study, Gazi Brains 2020 dataset, which contains normal and anomaly brain MRI data, was used to estimate the individual's age.

Curriculum learning was proposed to utilize the anomaly data in a better way. It was shown that using curriculum learning improved results compared to models who were trained on normal or both normal and anomaly data without curriculum learning. While using the entire dataset improved outcomes by nearly 20 % compared to using only the normal

patient data, curriculum learning improved the results by 20% further compared to classical training strategies.

This study shows that through the usage of curriculum learning, it is possible to utilize anomaly data with normal data in tasks unrelated to the anomaly itself.

The open datasets available online consist of much normal and anomaly data. In cases such as age detection, it is possible to utilize both the normal and anomaly data using curriculum learning. Using curriculum learning can increase the dataset size and ensure that the abnormalities cause as little bias as possible in tasks unrelated to the anomalies.

V. Future Work

In the scope of this work, further expansion on the concept of curriculum learning is planned. Since curriculum learning has had a considerable effect on the success of the model trained from this dataset, it is planned that the same concept will be applied to different datasets or amalgamation of datasets to see if a similar increase in performance can be observed.

The benchmarking of the effects of curriculum learning on various datasets will be presented and discussed. Furthermore, different sequences such as T1 or T2-weighted data will be added to examine the effect of curriculum learning further when different sequences are also part of training.

References

[1] L. K. Afshar and H. Sajedi, "Age prediction based on brain MRI images using extreme learning machine," in *2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, Bojnord, Iran, Apr. 2019, pp. 1–5. https://doi.org/10.1109/CFIS.2019.8692156

[2] H. Sajedi and N. Pardakhti, "Age prediction based on brain MRI image: A survey," *Journal of Medical Systems*, vol. 43, no. 8, Art no. 279, Jul. 2019. https://doi.org/10.1007/s10916-019-1401-7

[3] B. Wang, T. D. Pham, "MRI-based age prediction using hidden Markov models," *Journal of Neuroscience Methods*, vol. 199, no. 1, pp. 140–145, Jul. 2011. https://doi.org/10.1016/j.jneumeth.2011.04.022

[4] J. De Tobel, E. Hillewig, M. B. de Haas, B. Van Eeckhout, S. Fieuws, P. W. Thevissen, and K. L. Verstraete, "Forensic age estimation based on T1 SE and VIBE wrist MRI: Do a one-fits-all staging technique and age estimation model apply?" *European radiology*, vol. 29, no. 6, pp. 2924–2935, Jan. 2019. https://doi.org/10.1007/s00330-018-5944-7

[5] D. Štern, C. Payer, N. Giuliani, and M. Urschler, "Automatic age estimation and majority age classification from multi-factorial MRI data," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1392–1403, Sep. 2018. https://doi.org/10.1109/JBHI.2018.2869606

[6] S. Tangmose, K. E. Jensen, C. Villa, and N. Lynnerup, "Forensic age estimation from the clavicle using 1.0 T MRI – preliminary results", *Forensic Science International*, vol. 234, pp. 7–12, Jan. 2014. https://doi.org/10.1016/j.forsciint.2013.10.027

[7] J. A. Kra¨mer, S. Schmidt, K.-U. Ju¨ rgens, M. Lentschig, A. Schmeling, and V. Vieth, "Forensic age estimation in living individuals using 3.0 T MRI of the distal femur," *International Journal of Legal Medicine*, vol. 128, no. 3, pp. 509–514, Feb. 2014. https://doi.org/10.1007/s00414-014-0967-3

[8] B. Neumayer, M. Schloegl, C. Payer, T. Widek, S. Tschauner, T. Ehammer, R. Stollberger, and M. Urschler, "Reducing acquisition time for MRI- based forensic age estimation," *Scientific Reports*, vol. 8, Art no. 2063, pp. 1–9, 2018. https://doi.org/10.1038/s41598-018-20475-1

[9] D. Štern, C. Payer, V. Lepetit, M. Urschler, "Automated age estimation from hand MRI volumes using deep learning," in *International conference on medical image computing and computer-assisted intervention*, *Lecture Notes in Computer Science*, vol 9901, Springer, 2016, pp. 194–202. https://doi.org/10.1007/978-3-319-46723-8_23

[10] F. Fan, K. Zhang, Z. Peng, J.-H. Cui, N. Hu, Z.-H. Deng, "Forensic age estimation of living persons from the knee: comparison of MRI with radiographs," *Forensic Science International*, vol. 268, pp. 145–150, Nov. 2016. https://doi.org/10.1016/j.forsciint.2016.10.002

[11] B. Neumayer *et al.*, "The four-minute approach revisited: accelerating MRI-based multi-factorial age estimation," *International Journal of Legal Medicine*, vol. 134, pp. 1475–1485, Dec. 2019. https://doi.org/10.1007/s00414-019-02231-w

[12] E. Hillewig *et al.*, "Magnetic resonance imaging of the sternal extremity of the clavicle in forensic age estimation: towards more sound age estimates," *International Journal of Legal Medicine*, vol. 127, no. 3, pp. 677–689, Dec. 2013. https://doi.org/10.1007/s00414-012-0798-z

[13] Y. Guo *et al.*, "Dental age estimation in living individuals using 3.0 T MRI of lower third molars," International Journal of Legal Medicine, vol. 129, no. 6, pp. 1265–1270, Aug. 2015. https://doi.org/10.1007/s00414-015-1238-7

[14] P. Tramini, B. Bonnet, R. Sabatier, and L. Maury, "A method of age estimation using raman microspectrometry imaging of the human dentin," *Forensic Science International*, vol. 118, no. 1, pp. 1–9, Apr. 2001. https://doi.org/10.1016/S0379-0738(00)00352-2

[15] D. J. Madden, W. L. Whiting, S. A. Huettel, L. E. White, J. R. MacFall, and J. M. Provenzale, "Diffusion tensor imaging of adult age differences in cerebral white matter: Relation to response time," *NeuroImage*, vol. 21, no. 3, pp. 1174–1181, Mar. 2004. https://doi.org/10.1016/j.neuroimage.2003.11.004

[16] D. P. Varikuti *et al.*, "Evaluation of non-negative matrix factorization of grey matter in age prediction," *NeuroImage*, vol. 173, pp. 394–410, June 2018. https://doi.org/10.1016/j.neuroimage.2018.03.007

[17] L. Su, L. Wang, and D. Hu, "Predicting the age of healthy adults from structural MRI by sparse representation," in *International Conference on Intelligent Science and Intelligent Data Engineering*, *Lecture Notes in Computer Science*, vol 7751, Springer, 2012, pp. 271–279. https://doi.org/10.1007/978-3-642-36669-7_34

[18] P. Lam, A. H. Zhu, I. B. Gari, N. Jahanshad, and P. M. Thompson, "3D grid-attention networks for interpretable age and Alzheimer's disease prediction from structural MRI," *arXiv*, Art no. 2011.09115, 2020.

[19] I. Beheshti, S. Mishra, D. Sone, P. Khanna, and H. Matsuda, "T1-weighted MRI-driven brain age estimation in Alzheimer's disease and Parkinson's disease," *Aging and Disease*, vol. 11, no. 3, pp. 618–628, 2020. https://doi.org/10.14336/AD.2019.0617

[20] M. Ueda, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "An age estimation method using 3D-CNN from brain MRI images," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, Apr. 2019, pp. 380–383. https://doi.org/10.1109/ISBI.2019.8759392

[21] T.-W. Huang *et al.*, "Age estimation from brain MRI images using deep learning," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, Australia, Apr. 2017, pp. 849–852. https://doi.org/10.1109/ISBI.2017.7950650

[22] K. Armanious *et al.*, "Age-net: An MRI-based iterative framework for brain biological age estimation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 7, pp. 1778–1791, July 2021. https://doi.org/10.1109/TMI.2021.3066857

[23] K. Franke *et al.*, "Estimating the age of healthy subjects from T_1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *NeuroImage*, vol. 50, no. 3, pp. 883–892, Apr. 2010. https://doi.org/10.1016/j.neuroimage.2010.01.005

[24] C.-L. Chen *et al.*, "Generalization of diffusion magnetic resonance imaging–based brain age prediction model through transfer learning," *NeuroImage*, vol. 217, Art no. 116831, Aug. 2020. https://doi.org/10.1016/j.neuroimage.2020.116831

[25] L. Lin, G. Zhang, J. Wang, M. Tian, and S. Wu, "Utilizing transfer learning of pre-trained AlexNet and relevance vector machine for regression for predicting healthy older adult's brain age from structural MRI," *Multimedia Tools and Applications*, vol. 80, pp. 24719–24735, Apr. 2021. https://doi.org/10.1007/s11042-020-10377-8

[26] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning*, June 2009, pp. 41–48. https://doi.org/10.1145/1553374.1553380

[27] W. Wang, T. Ishikawa, and H. Watanabe, "Facial age estimation by curriculum learning," in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, Kobe, Japan, June 2020, pp. 138–139. https://doi.org/10.1109/GCCE50665.2020.9291929

[28] K. Li, J. Xing, W. Hu, and S. J. Maybank, "D2c: Deep cumulatively and comparatively learning for human age estimation," Pattern Recognition, vol. 66, pp. 95–105, June 2017. https://doi.org/10.1016/j.patcog.2017.01.007
[29] J. Kim, W. Bae, K.-H. Jung, I.-S. Song, Development and validation of deep learning-based algorithms for the estimation of chronological age using panoramic dental x-ray images (2019).
[30] Gazi Brains 2020 Datasetdoi:10.7303/syn22159468.
[31] S. Ji, W. Xu, M. Yang, K. Yu, 3d convolutional neural networks for human action recognition, IEEE transactions on pattern analysis and machine intelligence 35 (1) (2012) 221–231.



Alican Aşan recieved his B. Sc. degree in Control and Automation Engineering from Istanbul Technical University, Istanbul, Turkey, in 2017. He worked as a Machine Learning Research Engineer in Digital Transformation Office, Presidency of the Republic of Turkey for 3 years. He is currently a Computer Vision Engineer in Envision, a startup developing smart glasses to help visually impaired people with various tasks based in the Netherlands. His research interests mainly consist of computer vision, medical image processing and deep learning.
E-mail: alican@letsenvision.com
ORCID iD: https://orcid.org/0000-0002-9971-2678



Ramazan Terzi received his Bachelor of Science (B. Sc.), Master of Science (M. Sc.) degrees in computer engineering from Erciyes University, Kayseri, Turkey in 2010 and, 2013, respectively. He received the Ph. D. degree in Gazi University in Ankara, Turkey in 2019. He is currently a faculty member of the Department of Computer Engineering, Amasya University, Amasya, Turkey. His recent research interests focus on deep learning and medical image processing.
E-mail: ramazan.terzi@amasya.edu.tr
ORCID iD: https://orcid.org/0000-0003-2345-8666



Nuh Azginoglu received his Bachelor of Science (B. Sc.), Master of Science (M. Sc.) and the Ph. D. degrees in computer engineering from Erciyes University, Kayseri, Turkey in 2010, 2013, and 2019, respectively. He is currently a faculty member of the Department of Computer Engineering, Kayseri University, Kayseri, Turkey. His recent research interests focus on bioinformatics, data mining and deep learning.
E-mail: nuhazginoglu@kayseri.edu.tr
ORCID iD: https://orcid.org/0000-0002-4074-7366