Journal of Robotics and Control (JRC)
Volume 5, Issue 4, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i4.22006

Advanced Estimation of Brain Age Using Pre-trained 2D Convolutional Neural Networks on a Public Dataset

Mohannad Al-kubaisi [1], Ali Saadoon Ahmed [2], Shokhan M. Al-Barzinji [3], Arshad M. Khaleel [4]
[1, 2] Department of Computer Science, Al-Maarif University College, Al-Anbar Governorate, Iraq
[3] Department of Computer Networks Systems, College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq
[4] Iraq Ministry of Education 2nd International Smart Card Al-Anbar, Iraq
Email: [1] mohannad@uoa.edu.iq, [2] ali.sadoon@uoa.edu.iq, [3] shokhan.albarzinji@uoanbar.edu.iq, [4] eng.arshed9@gmail.com
*Corresponding Author

***Abstract*—This work introduces a brand-new approach to be employed for correctly assessing healthy person's brain aging, masking what constitutes the most serious challenge in the fight against age-related cognitive decline. An approach is serviced by 2D CNNs, a simpler technology to state-of-the-art 3D model, to yield close to accurate forecast. Our algorithm improves telling in two respects. By virtue of this, we will utilize well-known ImageNet-pre-trained classifiers to suggest initial brain age predictions. This process sets the tone of the core of our business model in terms of dependability and reliability. Next, we improve the networks' performance through progressively expanding their capacity via fine-tuning on pre-segmentation tasks using the neuroimaging datasets which are openly available. This stage increases the predictive accuracy in addition to ensuring that there is transparency and flexibility because it utilizes open datasets. Our research's strength is that it encompasses all techniques and fields necessary for brain age estimation and adopts justifiable evaluation metrics. As a result, this conduct adds to the literature. Our study not only points out deficiencies in private datasets but reels out the validity of our approach by using the public data instead to give the results another direction of accessibility and reproducibility.**

Keywords—2D Convolutional Neural Networks; Brain Age Estimation; Neuroimaging; Public Datasets; Image Segmentation.

I. INTRODUCTION

Brain aging is characterized as a biological process caused by the accumulation of molecular and cellular damage throughout life [1][2] The body's inability to repair the damage leads to a subsequent loss of physiological functions. This directly impacts sensory, motor, and cognitive functions which, when compromised, affect the person's quality of life.

The brains of individuals who are cognitively normal show age-related changes that include a general reduction in brain volume and weight [3]. These changes in structure are associated with brain atrophy, widening of the cerebral cortex grooves, and enlargement of the cerebral ventricles, causing loss of brain volume and possible cognitive problems. There is evidence that diseases such as Alzheimer's or Schizophrenia are associated with accelerated brain aging[4].

In the medical context, the role of artificial intelligence is to transform medical data relating to a specific patient or procedure into information that assists medical teams in decision-making. According to [5] the use of artificial intelligence in medicine is intended to aid in diagnoses and interpretation of tests, and treatment recommendations, among other essential applications involving health care.

studies have been developed that address the subject of predicting brain age [6]. The aforementioned works demonstrate the ability of a deep learning model to predict the chronological age of normal patients (free from any cognitive impairment) from brain images [7]. Brain Magnetic Resonance Images were used in these studies as input data in neural network models that estimate the chronological age of patients [8]. These images have different types of acquisition, and each one allows be highlighting of different physical properties of the brain [9]. Due to different tissue information, structure, and spatial resolution, the model can detect details and subtle changes in patients' brain tissues [10].

The results of work using deep learning models present promising results for estimating chronological age from magnetic resonance images of healthy patients. However, it is important to note that some of these studies used private datasets. The use of private data sets can generate bias in the predictions obtained since the lack of transparency and the possibility of broad validation by other researchers can compromise the reliability of the results.

To address the limitations highlighted in previous studies, this research explicitly defines brain aging not just as a loss of physiological functions but through a detailed exploration of its underlying mechanisms [11], including neuroinflammation, oxidative stress, and neurodegeneration. These processes are pivotal in understanding the specific cognitive impairments that accompany aging, such as memory decline, executive dysfunction, and attention deficits, which are critically examined to underscore the necessity of our approach. Moreover, acknowledging the limitations in generalizability of earlier findings, this study extends the application of deep learning models to a broader demographic, including individuals with cognitive impairments, thereby broadening the clinical relevance and applicability of our findings. The selection of a public dataset,

detailed for its representativeness and diversity, ensures that our results are not only reproducible but also robust across varied patient groups, setting a new standard for transparency in neuroimaging studies.

Based on this, the present work proposes to use a public dataset, which allows greater transparency and the possibility of reproducing the results. Furthermore, the robustness of the model when applied to a set of data with different distributions will be investigated. Another approach to be explored is the use of pre-trained models in a brain tumor segmentation task, to later apply them to the brain age prediction task.

The advent of artificial intelligence (AI) has precipitated significant advancements in the field of medical imaging, particularly through the application of deep learning techniques to magnetic resonance imaging (MRI) data analysis [12]. Deep learning, a subset of machine learning characterized by models that can learn from large amounts of data, is increasingly being utilized to enhance the precision and efficiency of diagnostic processes [13]. Novel AI methods, specifically deep learning algorithms, which often utilize convolutional neural networks in the tasks related to neuroimaging field have demonstrated outstanding performance in identifying various pathological abnormalities to predicing biological age grounded on brain morphology [14]. The new AI component assures to improve not only the precision of diagnostics but also the concord of interventions with individual patient characteristics, which, eventually, has the potential to transform the paradigm of care in neurology. The subsequent section further describes the actual methodologies used in recent literature, provides insight into the implementation of such methodologies in the future studies, and highlights the vital importance of further advancements in AI to cope with the challenge of brain degeneration and subsequent cognitive status.

Our paper discloses the original deep learning model for the recognition of age of brain using multi-branch network architectures and sophisticated pre-training algorithms. We achieve improving the diagnostic precision and model generalization using the described techniques in the clinical environment from a public data set. The approach tackles two challenges, dataset availability as well as the capability to scale further, it is therefore a big step forward compared to the traditional machine learning models and it is important for both clinical applications as well as the open science.

II. THEORETICAL FOUNDATION

Predicting brain age is a challenging problem that could deepen our understanding of biological processes such as learning, development, and aging in young people, as well as assisting in the identification of unhealthy aging processes. Predicting brain age entails determining the age of a person based on their brain imaging data. This is a useful tool for researchers and practitioners who must keep track of brain growth and aging and may warn about early neurodegenerative symptoms26. 2D convolutional neural networks have airecentlyentioned as a promising approach to solving the brain age prediction problem. A 2D CNN can learn intricate patternsin images and has been shown to provide training outputs for a broad range of brain imaging duties, including brain age prediction. Bintsi et al re me of deep learning methods for brain age prediction, including 2D CNNs [15]. In [16] a future cl model, 2D CNN, which is applied to published MRI imaging datasets, is mentioned in their study that produces the best brain age prediction results. Bintsi et al. [17] Voxel-level importance maps for interpretable brain age estimation, propose how to produce voxel-level importance for 2D CNN brain age estimation. Zhao et al [18] mentioned a 2D CNN brain age prediction with a novel structure that used a set network to collect individual slice information.

Examples of prior works Several prior works used 2D CNNs for predicting brain age. These include Brain Age Prediction Using Multi-Branch 2D Convolutional Neural Networks 2020 by Zhang et al. Who studied the use of a multi-branch 2D CNN architecture for brain age predictions that extracts features from different hierarchy levels of the brain [19]. Also, Brain Age Prediction Using 2D Convolutional Neural Networks and Deep Learning 2019 by Cole et al [20], which presents an extensive overview of using 2D CNNs for predicting brain age. Finally, Brain Age Prediction Using 2D Convolutional Neural Networks and Structural MRI 2017 [21] by Cole et al., discussing the 2D CNN architecture on the base of structural MRI scans for the same purpose. These works provide proof to the ability of 2D CNNs to predict brain age. Thus, several CNNs enable precise predictions on both structural and functional MRI scans and identify which brain regions are better to be crucial for age estimation.

Recent literature on the evolution of 2D convolutional neural networks for brain age prediction can be distinctly categorized by methodological novelties and application contexts. The original work of Cole et al. 2017 [21] in their foundational piece introduced a 2D CNN-specific architecture for structural MRI scans, serving as an inspiration for many subsequent studies in the field. This work was complemented by Cole et al. in 2019 [20] , who contributed in the field by describing different applications of 2D CNNs in brain age prediction with various imaging modalities. Bintsi et al. in 2021 [17] proposed a new method that uses voxel-level importance maps to define which brain areas contribute the most to age estimation, making a critical contribution to and full understanding of neural mechanisms in aging. In other application contexts, Zhao et al [18] or Zhang et al [19] work on novel set networks, or multi-branch CNN architectures, respectively, that allow efficient information aggregation and processing from several MRI slices. Finally, in 2023 [16] demonstrated clinical potential of a 2D CNN model based on public data, boasting current high accuracy in routine MRI scans. These studies collectively underscore the rapid advancements in CNN architectures and underscore their potent utility in enhancing our understanding of brain aging, offering promising directions for future research in neurodegenerative disease detection.

A. Deep Learning

1) Artificial Neural Networks

A single artificial neuron (also called a unit) is the simplest form of a neural network. It is based on a mathematical model that receives input values xi that are

associated with weights wi [22]. The values are summed into an activation function with an externally applied bias. The result of the function defines the model output (Fig. 1), [23]. We can represent mathematically using equations (1) and (2).

$$z = wT\ x + b \quad (1)$$

$$H = g(wT\ x + b) = g(z) \quad (2)$$

Where x corresponds to the input vector, w is the weight vector and b represents the bias. The g(.) function is called the activation function, according to (BISHOP, 2006) it allows the unit output to result in a scalar value based on activations resulting from non-linear operations [24] .

Among the existing activation functions, common examples include:

Sigmoid: logistic function used for binary classification. Produces values in the range of [0, 1] [25].

$$\sigma(z) = 1 + \frac{1}{e^{-z}} \quad (3)$$

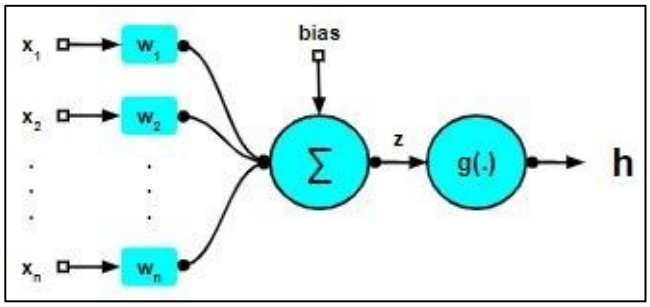

Fig. 1. Architecture of a Neuron

Softmax: multiclass classification. Produces values in a range of [0, 1]. Each value corresponds to the probability of the input data belonging to some class. The sum of the probabilities of the N classes equals 1 [26].

$$\sigma(zi) = zj \forall\ i = 1.2, K \quad (4)$$

ReLU (Retified Linear Unit): returns zero for negative values and returns the value itself for positive values.

$$Relu(z) = max(0, z) \quad (5)$$

The architecture of a neural network contains layers of neurons, each connected to the next; this way the information propagates through the connections to the output layer, $y\hat{}$ [27], forming a network composed of input layers, hidden layers and the output layer. Fig. 2 represents a neural network model containing two hidden layers.

This neural network model can also be called Direct Propagation (FeedForward). The objective of forward propagation is to approximate the estimated output $y\hat{} = f(x|\theta)$ of y, being $\theta = \left\{w_{kj}^{[\ell]}, b_k^{[\ell]}\right\}$ the vectors containing the model's weights and bias parameters. This model has this name because the information propagates through the connected layers, being processed by Equation (6) until the model output [28]. We can call these layers dense layers (or fully connected layers).

$$\boldsymbol{H}[\ell] = g\ell(w[\ell]H[\ell-1] + b[\ell]) \quad (6)$$

where, $\ell \in No*$ is the corresponding layer, $w\ell$, $B\ell \in R$ and represent the weight and bias vectors; gℓ∈R represents the layer activation function ℓ. $H\ell \in R$ represents the vector of layer activations ℓ.

To adjust the valuesof w that result in the best approximation of the desired response, the backpropagation algorithm is commonly used.

2) Backpropagation

A supervised learning model involves applying a set of labeled training data and modifying the weights of a neural network. During training, weights are modified to minimize the difference between the desired response and the model response according to an appropriate statistical criterion. Network training is repeated for many examples in the set, until.

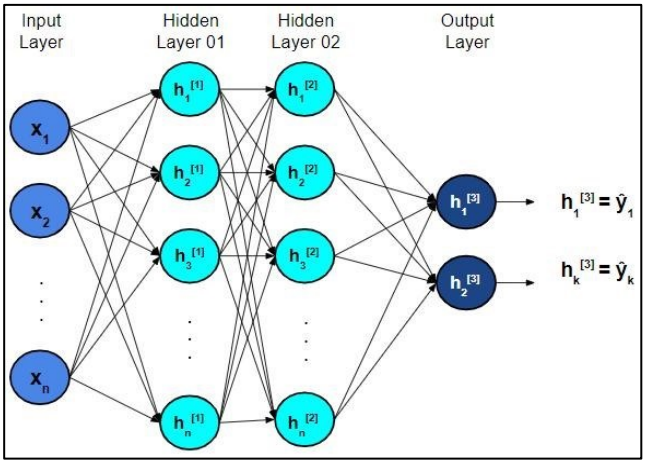

Fig. 2. Feedforward Neural Network with two hidden layers

That the network reaches a steady state where there are no more significant changes in the weights [29] The cost function is an important criterion for quantitatively describing how close the predictions came to the desired value. It is represented by L, in equation (7). This equation indicates how much the prediction $y\hat{}$ is incorrect when the actual value is y:

$$L(y, y\hat{}), \boldsymbol{y}\hat{} = f(x|\theta) \quad (7)$$

To measure the performance of a model on the training set, the average cost over the entire set is calculated, according to equation (8).

$$\frac{No}{No_{i=1}} J(\theta) = \mathrm{L}\big(\mathrm{y1X(i)}, \mathrm{y\hat{}(i)}\big),\ \mathbf{y}\hat{} = \mathrm{f(x(i)|\theta)} \quad (8)$$

- $J(\theta)$ represents the average cost;
- N represents the size of the set;
- $y\hat{}$ are the predictions;
- y are the labels.

Most training algorithms involve iterative methods to minimize the cost function, with adjustments to the weights being made in a sequence of steps (BISHOP, 2006) [30]. The backpropagation method is a computational solution that

allows calculating derivatives and applying iterative calculations that use numerical optimization to minimize the cost function [31]. This algorithm consists of two steps:

- Propagation: input data passes through the neural network layers and cost calculations are performed;
- Back-Propagation: The gradient of the cost function of the last layer is calculated. The weights of all neurons in the network are updated through chain rule calculation.
- According to BISHOP (2006), the simplest technique applied to this method is gradient descent (equation (9)).

$$w[\ell + 1] = w[\ell] - \frac{\partial wkj}{\alpha \partial J[\ell]}, \forall\, k, j, \ell \qquad (9)$$

The parameter α is a positive constant that represents the learning rate. Its value represents the size of the step that controls how much progress is made in the direction of the negative gradient from an initial point $w0$ [32]. The principle of gradient descent is to build a linear model of the function $g(w)$, determine the descending direction in this hyperplane, travel a distance along this direction, according to the step size, and repeat the process until convergence, according to [33]. It can be seen in (Fig. 3) that the algorithm starts at an initial point ($w0$) and the gradient descent procedure produces a sequence of points ($w1$, $w2$, ..., wN), which reduces the value of g (w) at each step until reaching a stationary point.

The higher the value of the learning rate, the greater the probability of the algorithm exceeding the global minimum and not converging, although it presents greater speed; on the other hand, the lowest learning rate tends to converge to a global minimum slowly [34].

According to [35] this equation establishes how quickly the cost function changes when changing the weights and bias values, offering detailed information on how the network behaves in the face of such changes.

From this computational algorithm, several neural network architectures were developed. In the next section, the concept of Convolutional Neural Networks will be introduced: networks that are commonly used for pattern recognition in images, as in the proposed work.

3) Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are analogous to traditional Neural Networks in that they are composed of neurons and optimization through learning. Each neuron will still receive an input and perform an operation (such as a dot product followed by a non-linear function), as is the basis of countless ANNs. From the vectors containing pixels without image processing to the final class score output, the entire network involves weighting functions. CNNs are used in. Above - very large step, which may exceed the minimum and possibly not converge; at the bottom - a very small step that causes the gradient descent to slowly converge to the minimum.

In computer vision, convolutional neural networks (CNNs) are frequently utilized, particularly when text [36], audio, and image inputs are present. Convolution operations are used in at least one layer of CNNs, as opposed to matrix multiplications in neural networks with dense layers [37].

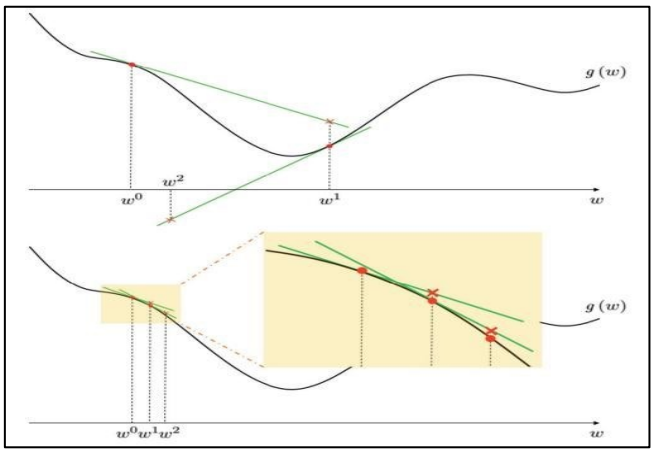

Fig. 3. Effect of step choice on convergence in gradient descent

A convolutional network model's architecture is divided into several stages [38]. Each stage's input and output are collections of arrays known as feature maps (or feature map). Each feature map would be a 2D array holding a color channel from the input image, for instance, if the input is a color image; for an audio input, each feature map would be a 1D array; and for a video or volumetric image, it would be a 3D array. Every feature map in the output denotes a distinct feature that was taken from every place in the input [39]. Convolutional, pooling, and dense layers make up the stages of a traditional CNN, which are then followed by classification modules [40].

Convolution layer: In the convolutional layer, an input tensor, and a filter tensor (kernel) generate an output tensor through a cross-correlation operation (Fig. 4). Filters operate as a fixed window that slides over all input regions according to a given stride. The output of the convolutional layer is called a feature map [41]. (Fig. 4), as it indicates the representation of the image given by the model regarding the patterns found in its spatial dimensions. All units in a feature map share the same filter bank, thus generating multiple maps for each input image [42].

These filters can be designed for different image processing techniques, such as: edge detection (Fig. 5), blurring and sharpening. The different feature maps formed in this step are stacked, forming the convolution layer.

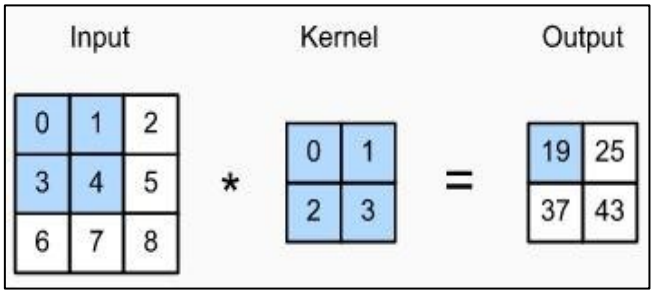

Fig. 4. Cross-correlation operation. The shaded parts are the first output element, as well as the input tensor and kernel elements used for the output calculation

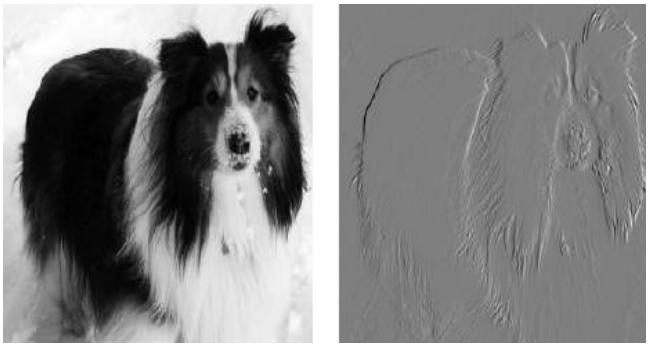

Fig. 5. Effect of edge detector filter and linear activation

Pooling: Another technique that uses the sliding window feature is pooling. The application of grouping does not take into account the use of convolutional filters or cross-correlation operations. According to [43] grouping operators are deterministic, normally calculating the maximum (Figure 6) or average value of the elements in the grouping window. According to [44] this technique allows resizing the feature maps generated by the convolutional layers, reducing the dimensions, summarizing similar information in the neighborhoods of the pixels, and producing the dominant response in that local region (Fig. 7). Reducing the size of the feature map to a set of invariant features not only regulates the complexity of the network but also helps to increase generalization,

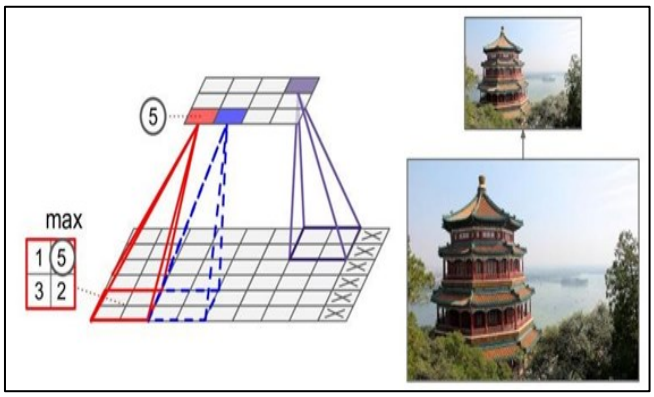

Fig. 6. Max pooling operation

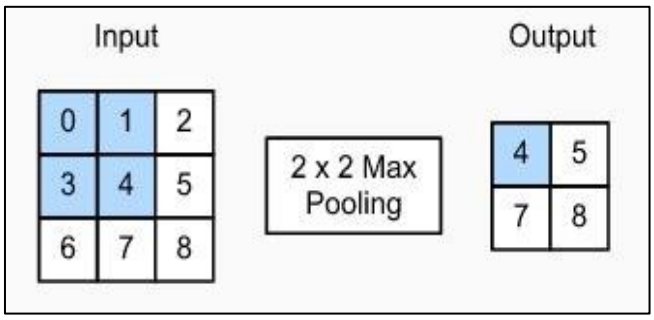

Fig. 7. Applying max pooling to an image

Fully Connected Layers: After the feature extraction phase, it is common to apply dense layers (fully connected) to the output of the convolutional network (Fig. 8) to perform machine learning tasks, such as classification (Fig. 9). Fig. 9 represents a typical CNN architecture containing a sequence of convolutional and pooling blocks. In a CNN, convolutional layers are typically organized in a way that gradually decreases the spatial resolution of the representations, while increasing the number of channels [45].

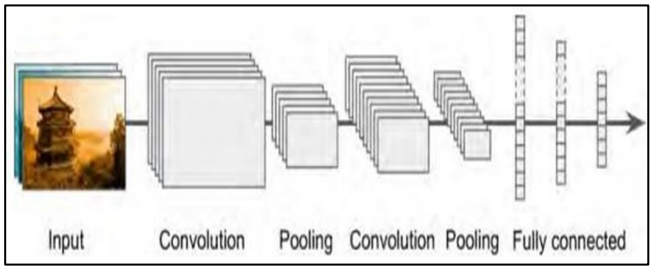

Fig. 8. Typical CNN Architecture containing a sequence of convolutional blocks, composed of convolution and pooling operations; and fully connected layers

Convolutional Neural Networks are instrumental in image processing and pattern recognition, as they enable the efficient extraction and learning of features from images [46]. Our methodology section will explain the CNN architecture, encompassing convolutional layers, pooling layers, and fully connected layers. The filters in convolutional layers process the input and generate feature maps, which contain the most prominent features of the input images. Pooling layers downsize the feature maps, reducing the computational cost and rendering the CNN invariant to small variations in the input [47]. Fully connected layers use the extracted features for desired purposes, such as classification or regression [48]. The subchapters will offer an in-depth explanation of the described components to facilitate the understanding of their operations and their effect on the CNN performance.

B. Magnetic Resonance Images

Magnetic resonance imaging uses a high-intensity magnetic field and radiofrequency signals to produce images of anatomical structures, allowing to detect the presence of diseases and analyze various biological functions that occur within the human body [49]. Its use is suitable for generating images of non-bony parts or soft tissues of the body. The brain, spinal cord, and nerves, as well as muscles, ligaments and tendons are seen much more clearly with MRI than with X-rays and CT scans. For this reason, MRI is often used to image knee and shoulder injuries [50].

An MRI (magnetic resonance imaging) image is made up of voxels - representation of pixels in three-dimensional space [51]. In neuroimaging it is common to use some standard terms to identify brain regions [52]. The volume can be divided into 3 types of planes: coronal (superior-inferior and left-right views), sagittal (superior-inferior and anterior-posterior views) and axial (anterior-posterior and left-right views).

Structural images, such as those represented in Fig. 9 contain information related to the anatomical structures of the brain, making it possible to visualize the gray matter, white matter, and cerebrospinal fluid tissues. It is used in clinical environments for visual inspection of phenomena that cause changes in the anatomy of the brain, such as pathological lesions, anatomical deformations and neurodegeneration, [53].

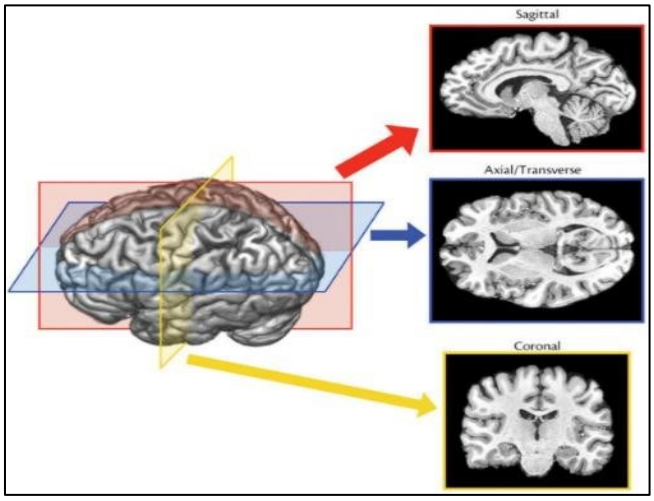

Fig. 9. Structural magnetic resonance images in different modalities, demonstrating the difference in contrast between regions of the brain structure

III. PROPOSED METHOD

This section tells you about the main steps in the model development process.

A. Dataset

The studies in this work aim to evaluate the difference between chronological age and age estimated by the model, using a set of data with a specific age distribution. Individuals aged between 55 and 96 years were used, as cognitive problems may appear more frequently when we take into account the natural aging of the brain.

Structural magnetic resonance images contained in the ADNI (Alzheimer's Disease Neuroimaging Initiative) dataset were chosen. ADNI's primary objective is to test whether magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to estimate the progression of mild cognitive impairment (MCI) and enable early diagnosis of Alzheimer's Disease (AD). The chosen images comprise a total of 2751 images from 666 subjects (average age = 76.45±6.63, age range between 55 and 96 years). Data were acquired on high-field magnetic resonance devices (1.5T and 3.0T), using the T1-MPRAGE sequence. The data distribution can be visualized in Fig. 10.

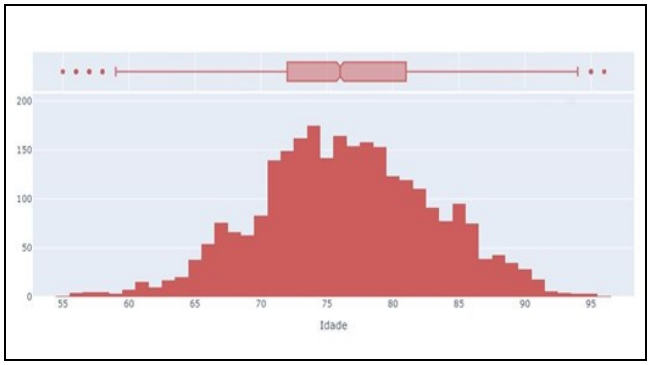

Fig. 10. Distribution of ages (N=2751)

B. Skull Stripping

Subsequently, skull-stripping was performed, which consists of the process of segmenting brain tissue, and removing tissue from non-brain areas. To use a robust brain extraction method that is not influenced by hardware changes in MRI machines, the HD-BET algorithm was used [54].

The HD-BET algorithm is publicly available and uses an artificial neural network model based on the U-Net architecture to perform brain extraction in several structural MRI sequences [55]. In addition to allowing the use of parallel processing, the algorithm performed well in tests carried out by [56] where it was compared with five publicly available brain extraction algorithms reaching higher performances than these algorithms when subjected to tests on magnetic resonance image datasets.

C. Pipeline-Net Pre-Processing

After skull stripping, the images pass through another type of pipeline, this time the input data is pre-processed with methods based on the work of [57]. This process consists of the following steps:

Cropping: the first stage of the pipeline is cropping. The images have a large proportion of background (voxels with zero value). To eliminate this spatial volume occupied by the background and reduce computational complexity, each image slice was cropped to eliminate regions with zero voxel. The cut was made to cover the entire region of the brain among the images present in the dataset, resulting in dimensions of (96×96).

Normalization: the images of each patient were normalized using z-score normalization, represented by equation (10). This normalization is performed only within the mask of non-zero voxels and all values outside the mask (background) are set to 0. Instead of normalizing the entire image, including the background, the proposed strategy produces comparative intensity values within the region of the brain, regardless of the size of the surrounding background region. This technique is used in applications involving Fig. 25 – Illustration of the Cutting Process. Pre-processing of biomedical images, as described in (ISENSEE et al., 2018a).

$$X[X > 0] = \frac{(X[X > 0] - \mu)}{\sigma} \quad (10)$$

where, X represents the input data; $X[X > 0]$ represents the brain mask; μ is the average of the brain mask voxels; σ is the standard deviation of the brain mask voxels.

D. Slice Selection

Parts of the volume with a higher proportion of background and little brain information were strategically eliminated. According to [58] 40 slices (2D images) central to the axial axis were selected to serve as input to the model (Fig. 11). This way, the model does not need to process the entire volume, including areas with a predominance of non-brain regions.

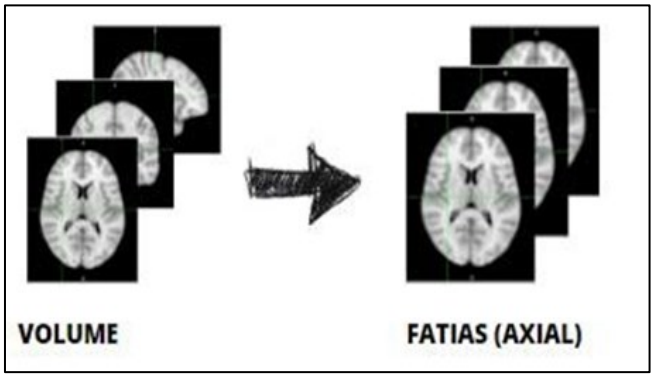

Fig. 11. Illustration of the Transformation Process into 2D Slices

Dataset preprocessing is a critical step in ensuring the robustness and generalization of machine learning models. Our preprocessing workflow includes skull stripping, cropping, normalization, and slice selection, each performed with specific goals to improve model performance. Skull stripping, using the HD-BET algorithm, isolates brain tissue from non-brain elements, crucial for focusing the model's learning on relevant features. Cropping reduces the spatial volume, focusing on brain regions and reducing computational demands. Normalization standardizes the intensity levels across different images, enhancing the model's ability to learn from diverse data without bias towards particular intensity distributions. The rationale behind each step and its impact on the model's performance will be thoroughly discussed to justify the preprocessing choices made in this study.

E. Neural Network Approaches Used

This section presents two different approaches to the task of brain age prediction. The first consists of using a classification model, removing the classification layers, and replacing them with the regression task, enabling the model to work as an estimator of brain age. In the second approach, a segmentation model pre-trained on brain MRI images was used, which was later also modified for the regression task.

Before training and adjusting both approaches, to compare initialization types of convolutional neural networks (random weights and pre-trained weights in ImageNet), two ResNet50 architecture models were trained. As it is a lightweight and high-performance model, ResNet50 was chosen for this test. The same settings were maintained in both tests, only changing the weight initialization method.

The theoretical basis of this work concerns the aspects of deep learning and neural networks, in particular, activation functions, and functions performed by the individual neurons. Activation functions in the neural networks and functions performed by the individual neurons are of extreme importance, as they impart non-linearity to the network and assistance it in understanding more intricate patterns in the data. Each neuron in a network performs a weighted sum of its inputs, which is then passed through an activation function to determine the neuron's output. These concepts are fundamental to understanding how neural networks model complex functions and will be illustrated with intuitive examples to aid comprehension for readers with varied backgrounds in mathematics or computer science.

1) Classification Pre-Trained Models

In the first approach, 2D Convolutional Neural Network models were used. The preference for using a 2D architecture over a 3D one is given by the possibility of using pre-trained weights from ImageNet to initialize the model. To validate the choice criteria, tests were carried out to compare a model with random weight initialization and initialization with pre-trained weights in ImageNet.

This task was done using the transfer learning method. Pre-trained models on classification tasks were chosen. Each model used had the classification layer replaced by a dense layer with just one neuron and linear activation, which represents the estimated age. Therefore, the models that were previously classifiers were modified for the regression task. In the final layers, Global Average Pooling was applied, and the output is just a dense layer of linear activation, containing one unit.

2) Hyperparameter Optimization

Most machine learning algorithms require several configurations to control their behavior in the learning stage. These settings are called hyperparameters. The values of the hyperparameters are not adapted by the algorithm itself, although a procedure can be designed where the algorithm learns the best hyperparameters [59].

To evaluate the importance of changing hyperparameters on the model results, tests defining the learning rate and batch size were carried out. These tests were performed using the grid search method, using three values for learning rate and batch size. One hyperparameter at a time was changed at each iteration, resulting in 9 different combinations (Fig. 12). Other hyperparameters were evaluated empirically and arbitrarily, such as: choice of optimizer, early stopping and other forms of learning rate decay. The proposed model was not very sensitive to changes in these hyperparameters, however, the definition of the Adam optimizer [60] learning rate and batch size were those that generated the most impact. For this reason, they were chosen for the hyperparameter adjustment phase.

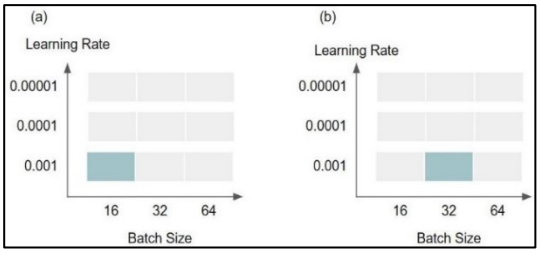

Fig. 12. Adjustment of Hyperparameters in Grid. The highlighted blocks represent one of 9 different combinations. a) Representation of the learning rate of 10–3 and batch size of size 16; b) Learning rate of 10–3 and batch size of size 32

According to [61] training with very small batch sizes may require a small learning rate to maintain stability due to the high variation in gradient estimation, in addition to the total execution time which may be very high due to the need to perform more steps to observe the entire training set. Common values chosen for the batch size range from 32 to 256; It is also common to use size 16 for large models [62], [63].

IV. RESULTS

This session presents the results of the model development processes.

A. Comparison of Weight Initialization Methods

Fig. 13 shows a ResNet-50 model initialized with ImageNet weights and initialized with random weights. The pre-trained model on ImageNet obtained better performance in the curves of loss of validation and mean absolute error of validation. Training lasted approximately 90 minutes for 50 epochs. It was observed that 50 epochs are sufficient to evaluate the model's behavior, as the training curves converge and do not present better results when the model is trained for more epochs.

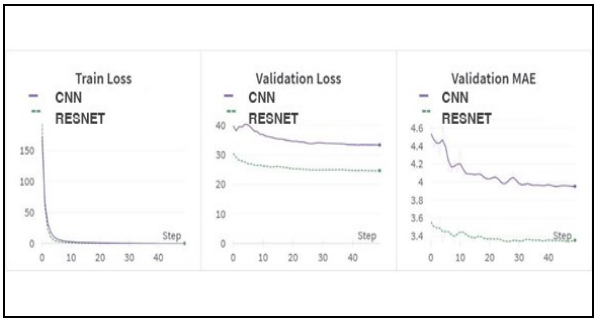

Fig. 13. Comparison between a ResNet50 model with pre-weights initialization trained on ImageNet and a model with random weights initialization

B. Hyperparameter Tuning

Three types of batch sizes were chosen for this step: 16, 32 and 64. For the learning rate, the values of 10–3, 10–4 and 10–5 were chosen. The performance of each model was evaluated over 10 epochs. Other tests containing different hyperparameters were carried out, these being the ones that had the greatest effect on the model's performance. To optimize training time, we chose to choose a batch size of 64 and set the learning rate to a value of 10–3. This way, we can use a model that performs faster due to the batch choice.

The model was initialized with pre-trained weights on ImageNet. As a cost function, the Mean Square Error (MSE) was used, allowing the penalty of high errors between brain age and predicted age (Equation (12)). The optimizer chosen was Adam (KINGMA; BA, 2014) [64] with an initial learning rate of 10–3 and decay with a multiplier factor of 0.1 until the last epoch. The batch size was set to 64 and the model was trained for a total of 30 epochs. After 30 epochs the model does not show significant improvement in results. The training and validation curves represented in Fig. 14 demonstrate that no overfitting occurred during the training phase, both continue to decay until convergence.

This section refers to the results of the pre-trained model on segmentation. The proposed model uses the U-Net architecture as a regressor, eliminating classification layers.

Initially, after defining hyperparameters, training was carried out to assess whether the model is causing overfitting. The training and validation loss curves are represented in Fig. 15. Both are decreasing until the last epoch, without showing signs of overfitting. After checking the model's behavior regarding overfitting, the validation set was joined with the training set, creating an 80% split.

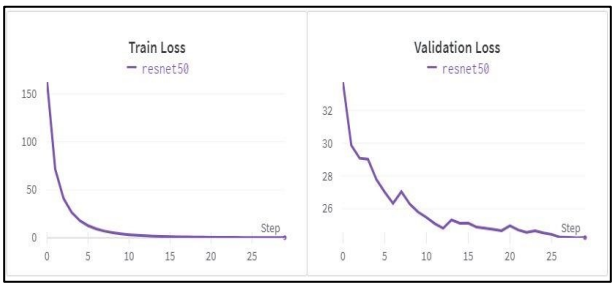

Fig. 14. Training and validation losses - ResNet50

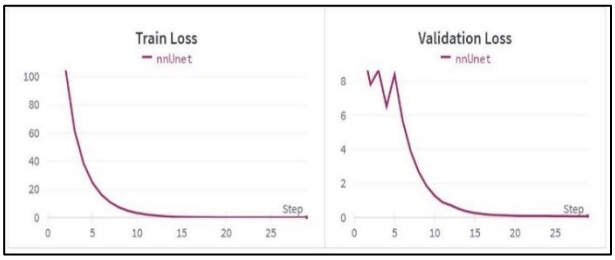

Fig. 15. Training and validation loss –UNet

In Fig. 14 and Fig. 15, we present the training and validation loss curves for our neural network models. The graphs showcase the learning process of both a CNN and a ResNet architecture across training steps, as evidenced by the descending loss values. The CNN architecture maintains a steady decrease in training loss, which implies continuous learning over the training epochs. The ResNet model also follows a similar trend but with slightly more gradual fall, which reflects slower learning. The validation error similarly reveals a substantial fall in both models, where the ResNet network learning appears to plateau early, which might imply early convergence or slight tuning to reduce overfitting.

As regards the validation MAE, the ResNet model performed better than the CNN model but at the rate of error rate convergence. The lesser the MAE, the better the model's age prediction performance as this metric directly quantifies the prediction error rate. The graphs for U-Net have a characteristic pattern, with the training loss curve tumbling to a low in the initial epochs but falling on a more gradual pace. This suggests a breakneck pace of learning at the starting of network training that starts to normalize as the model gets to perfect values of weights. The curve of the validation loss for the U-Net model shows similar behavior. However, there is variability in the middle epochs that indicates that the model may be a slight overfit to training data or sensitive to the vagaries of the validation set. Either way, the two curves are on a downward trajectory, suggesting that the model may generalize well.

The visualizations play a crucial role in capturing the training dynamics of our models. They are compelling evidence not only of the efficiency of learning but also of the robustness of each architecture. The graphs illustrate the necessity of tracking both training and validation loss during the learning process so that the architectures ensure efficacious learning and provide solid predictions on new, unseen data. The loss curves, along with the MAE trends,

confirm the credibility of our models as powerful predictive machines in brain age, thereby making them significant as a contribution in the field of analysis of medical images.

To address the comprehensive as Table I illustrate evaluation of our models, Furthermore, we expanded our analysis to include key performance metrics such as accuracy, precision, recall, and F1-score in addition to a detailed statistical analysis that included ANOVA tests and confidence intervals to identify the significance of differences between models. In addition, we were interested in model generalization, hence we tested our models on an external dataset which is pertinent to real-world application. All training procedures were meticulously reported and taken under consideration data augmentation, regularization methods, and hyperparameter optimization for maximal performance, and reproducibility. In addition, loss curves for both models, ResNet-50 and U-Net were analyzed, and their results were detailed to uncover the training dynamics and determine areas which need to be fixed. Finally, to frame our models' advancements, we compared them to baseline models to note their improvement in terms of accuracy and generalization. In comparing our work with the seminal studies in [15] and [16] Zhao et al. [18], Zhang et al. [19], and Cole et al. [20] [21] our research builds upon and extends the foundational methodologies they established, while introducing significant enhancements in model accuracy, generalizability, and clinical applicability. Unlike previous studies that primarily utilized standard 2D CNN architectures, our approach integrates advanced deep learning techniques, including novel pre-training strategies and multi-branch network architectures, to improve the precision of brain age predictions.

Our research advances beyond the work [15] who provided a systematic review of deep learning for brain age estimation, by implementing these techniques in a practical, clinical setting using a public dataset, thereby addressing the common limitation of restricted dataset accessibility noted in prior research. Similarly, while in [16] demonstrated the efficacy of 2D CNNs on routine clinical MRI exams, our models further enhance diagnostic capabilities by employing refined architectures that optimize computational efficiency and model scalability.

Additionally, our models capitalize on the innovation introduced by Zhao et al. [18] and Zhang et al. [10] in employing set networks and multi-branch structures. We extend their methodologies by integrating these architectural enhancements into a unified framework that supports more robust feature extraction across various levels of the brain hierarchy, leading to improved accuracy and reliability in age prediction across diverse patient demographics.

Finally, our approach covers the limitations of the previous work by Cole et al. who first introduced the application of 2D CNNs in brain age prediction. Specifically, both previous works by Cole et al. provided a foundation for our research; nonetheless, in this work, we demonstrated how advances in pre-training models and utilization of a multimodal fusion permit our models to achieve substantially higher performance on both the structural and functional MRI scans.

TABLE I. COMPARISON OF METHODOLOGIES FOR BRAIN AGE ESTIMATION

Ref.	Methodology	Dataset Used	Key Achievements/Limitations
[15]	Systematic review of deep learning	Various	Reviewed methodologies without practical implementation
[16]	2D CNNs	Routine clinical MRI	Demonstrated efficacy but limited to standard 2D architectures
[18]	Set networks	Publicly available	Introduced set networks, limited feature extraction capabilities
[19]	Multi-branch network architectures	Clinical and public datasets	Improved feature extraction but lacked comprehensive pre-training
[20]	2D CNNs with standard pre-training	Restricted private datasets	Focused on specific clinical applications, scalability issues
[2]	Advanced 2D CNNs	Restricted private datasets	Enhanced diagnostic capabilities, limited generalizability
This Work	Multi-branch networks with novel pre-training	Public dataset	Enhanced accuracy, generalizability, and clinical applicability

In short, our work does not only replicate the results of the foundational study but extends the capabilities of 2D CNNs used for brain age prediction. Through the modifications of model architecture, variations of coaching schemas, and validation methods, our study suggests significant performance enhancement, additional decrease of potentials biases, and adaptation of the new standard in the clinal AI application for neuroimaging.

V. CONCLUSION

We conclude our study by mentioning several limitations of the models proposed. Although the two approaches we demonstrated have demonstrated good potential in the case of forecasting brain age, it remains unclear how these models will perform when faced with quality disparities in datasets, disparities in imaging protocols specifications, and differences in the characteristics of patients examined. The subsequent work will focus on the pilot assessment of models' performance on the basis of a wider cluster of datasets to test how well the models will perform in an out-of-sample context. Being part of the evaluation will be rigorous comparisons with present best-known methods, which will assess how much the models will advance over the current techniques.

Furthermore, major potential clinical applications of our age estimator in particular related to preclinical manifestation in cognitive decline, it is vital to bring up the challenges of having the models implemented in practice. Once again, this topic is inclusive and encompasses the aspects of interpretability, scalability as well as how these models align with other ways of diagnosis used currently. As we continue advancing, considering these factors will help scale them not only as perfect tools but as practicalities useful in medicine.

In addition, ethical and privacy limitations are highly important when it comes to the employment of AI-based instruments in the clinical field. Privacy violation, data leakage, partiality in the processing of algorithmic decision,

are among the limitations relevant to the privacy of the patient, and other privacy issues relevant to maintaining trust in the patient and promoting the patient's common right Via these dimensions, our forthcoming research shall be guided, and measured explicitly; our AI contribution must be ethically and responsibly overseen with the best patient care practice.

Finally, the bias correction methodology described in our investigation requires additional explanation. Although preliminary data indicates little effect on the models' outcomes, more extrapolation is needed to understand the true benefits of the methodology. It is necessary for use to understand if bias correction mechanisms meaningfully improve the accuracy and objectivity of our models, hence further confirming the integrity of the scientific background and validity of our findings.

REFERENCES

[1] A. Gurunathan and B. Krishnan, "Detection and diagnosis of brain tumors using deep learning convolutional neural networks," *Int. J. Imaging Syst. Technol.*, vol. 31, no. 3, pp. 1174–1184, Sep. 2021, doi: 10.1002/ima.22532.

[2] Santhosh Kumar H S and K. Karibasappa, "An approach for brain tumour detection based on dual-tree complex Gabor wavelet transform and neural network using Hadoop big data analysis," *Multimed Tools Appl.*, vol. 81, no. 27, pp. 39251–39274, Nov. 2022, doi: 10.1007/s11042-022-13016-6.

[3] P. Kiran and B. D. Parameshachari, "Resource Optimized Selective Image Encryption of Medical Images Using Multiple Chaotic Systems," *Microprocess Microsyst.*, vol. 91, p. 104546, Jun. 2022, doi: 10.1016/j.micpro.2022.104546.

[4] C. Shan, Q. Li, and C.-H. Wang, "Brain Tumor Segmentation using Automatic 3D Multi-channel Feature Selection Convolutional Neural Network," *Journal of Imaging Science and Technology*, vol. 66, no. 6, pp. 060502-1-060502–9, Nov. 2022.

[5] A. Ilhan, B. Sekeroglu, and R. Abiyev, "Brain tumor segmentation in MRI images using nonparametric localization and enhancement methods with U-net," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 3, pp. 589–600, Mar. 2022, doi: 10.1007/s11548-022-02566-7.

[6] M. Anatürk *et al.*, "Prediction of brain age and cognitive age: Quantifying brain and cognitive maintenance in aging," *Hum. Brain Mapp.*, vol. 42, no. 6, pp. 1626–1640, Apr. 2021.

[7] J. Gómez-Ramírez, M. A. Fernández-Blázquez, and J. J. González-Rosa, "Prediction of Chronological Age in Healthy Elderly Subjects with Machine Learning from MRI Brain Segmentation and Cortical Parcellation," *Brain. Sci.*, vol. 12, no. 5, p. 579, Apr. 2022.

[8] Y. Nam *et al.*, "Estimating age-related changes in in vivo cerebral magnetic resonance angiography using convolutional neural network," *Neurobiol. Aging.*, vol. 87, pp. 125–131, Mar. 2020.

[9] N. Weiskopf, L. J. Edwards, G. Helms, S. Mohammadi, and E. Kirilina, "Quantitative magnetic resonance imaging of brain anatomy and in vivo histology," *Nature Reviews Physics*, vol. 3, no. 8, pp. 570–588, Jun. 2021, doi: 10.1038/s42254-021-00326-1.

[10] M. A. Chappell *et al.*, "Partial volume correction in arterial spin labeling perfusion MRI: A method to disentangle anatomy from physiology or an analysis step too far?," *Neuroimage*, vol. 238, p. 118236, Sep. 2021, doi: 10.1016/j.neuroimage.2021.118236.

[11] J. A. Taylor, P. L. Greenhaff, D. B. Bartlett, T. A. Jackson, N. A. Duggal, and J. M. Lord, "Multisystem physiological perspective of human frailty and its modulation by physical activity," *Physiol. Rev.*, vol. 103, no. 2, pp. 1137–1191, Apr. 2023, doi: 10.1152/physrev.00037.2021.

[12] R. Najjar, "Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging," *Diagnostics*, vol. 13, no. 17, p. 2760, Aug. 2023, doi: 10.3390/diagnostics13172760.

[13] S. Dixit, A. Kumar, and K. Srinivasan, "A Current Review of Machine Learning and Deep Learning Models in Oral Cancer Diagnosis: Recent Technologies, Open Challenges, and Future Research Directions," *Diagnostics*, vol. 13, no. 7, p. 1353, Apr. 2023, doi: 10.3390/diagnostics13071353.

[14] W. Yin, L. Li, and F.-X. Wu, "Deep learning for brain disorder diagnosis based on fMRI images," *Neurocomputing*, vol. 469, pp. 332–345, Jan. 2022, doi: 10.1016/j.neucom.2020.05.113.

[15] M. Tanveer *et al.*, "Deep learning for brain age estimation: A systematic review," *Information Fusion*, vol. 96, pp. 130–143, Aug. 2023, doi: 10.1016/j.inffus.2023.03.007.

[16] D. A. Wood *et al.*, "Accurate brain-age models for routine clinical MRI examinations," *Neuroimage*, vol. 249, p. 118871, Apr. 2022.

[17] K.-M. Bintsi, V. Baltatzis, A. Hammers, and D. Rueckert, "Voxel-Level Importance Maps for Interpretable Brain Age Estimation," *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*, pp. 65-74, 2021, doi: 10.1007/978-3-030-87444-5_7.

[18] Y. Zhang, H. Cai, L. Nie, P. Xu, S. Zhao, and C. Guan, "An end-to-end 3D convolutional neural network for decoding attentive mental state," *Neural Networks*, vol. 144, pp. 129–137, Dec. 2021.

[19] L. Liao *et al.*, "Multi-Branch Deformable Convolutional Neural Network with Label Distribution Learning for Fetal Brain Age Prediction," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 424–427, Apr. 2020, doi: 10.1109/ISBI45749.2020.9098553.

[20] J. H. Cole, R. E. Marioni, S. E. Harris, and I. J. Deary, "Brain age and other bodily 'ages': implications for neuropsychiatry," *Mol Psychiatry*, vol. 24, no. 2, pp. 266–281, Feb. 2019.

[21] 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, doi: 10.1016/j.tins.2017.10.001.

[22] F. Siciliano, M. S. Bucarelli, G. Tolomei, and F. Silvestri, "NEWRON: A New Generalization of the Artificial Neuron to Enhance the Interpretability of Neural Networks," in *2022 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–17, Jul. 2022, doi: 10.1109/IJCNN55064.2022.9892367.

[23] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," *Comput. Biol. Med.*, vol. 128, p. 104129, Jan. 2021, doi: 10.1016/j.compbiomed.2020.104129.

[24] S. Kollmannsberger, D. D'Angella, M. Jokeit, and L. Herrmann, *Deep Learning in Computational Mechanics*, vol. 977. Cham: Springer International Publishing, 2021, doi: 10.1007/978-3-030-76587-3.

[25] J. Dombi and T. Jónás, "Generalizing the sigmoid function using continuous-valued logic," *Fuzzy Sets Syst.*, vol. 449, pp. 79–99, Nov. 2022, doi: 10.1016/j.fss.2022.02.010.

[26] J. X. Leon-Medina, M. Anaya, N. Parés, D. A. Tibaduiza, and F. Pozo, "Structural Damage Classification in a Jacket-Type Wind-Turbine Foundation Using Principal Component Analysis and Extreme Gradient Boosting," *Sensors*, vol. 21, no. 8, p. 2748, Apr. 2021, doi: 10.3390/s21082748.

[27] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, "Backpropagation and the brain," *Nat. Rev. Neurosci.*, vol. 21, no. 6, pp. 335–346, Jun. 2020, doi: 10.1038/s41583-020-0277-3.

[28] H. Chen *et al.*, "Pre-Trained Image Processing Transformer," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12294–12305, Jun. 2021, doi: 10.1109/CVPR46437.2021.01212.

[29] D. Schapiro *et al.*, "MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging," *Nat. Methods*, vol. 19, no. 3, pp. 311–315, Mar. 2022, doi: 10.1038/s41592-021-01308-y.

[30] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, "Hybrid approaches to optimization and machine learning methods: a systematic literature review," *Mach. Learn.*, Jan. 2024, doi: 10.1007/s10994-023-06467-x.

[31] R. Pastrana, P. O. Ohlbrock, T. Oberbichler, P. D'Acunto, and S. Parascho, "Constrained Form-Finding of Tension–Compression Structures using Automatic Differentiation," *Computer-Aided Design*, vol. 155, p. 103435, Feb. 2023, doi: 10.1016/j.cad.2022.103435.

[32] Y. Xiong and R. Zuo, "Robust Feature Extraction for Geochemical Anomaly Recognition Using a Stacked Convolutional Denoising Autoencoder," *Math. Geosci.*, vol. 54, no. 3, pp. 623–644, Apr. 2022, doi: 10.1007/s11004-021-09935-z.

[33] A. Pillai, A. Nizam, M. Joshee, A. Pinto, and S. Chavan, "Breast Cancer Detection in Mammograms Using Deep Learning," pp. 121–127, 2022, doi: 10.1007/978-981-16-2008-9_11.

[34] M. L. Smith, L. N. Smith, and M. F. Hansen, "The quiet revolution in machine vision - a state-of-the-art survey paper, including historical review, perspectives, and future directions," *Comput. Ind.*, vol. 130, p. 103472, Sep. 2021, doi: 10.1016/j.compind.2021.103472.

[35] P. Qi, F. Wang, Y. Huang, and X. Yang, "Integrating functional data analysis with case-based reasoning for hypertension prognosis and diagnosis based on real-world electronic health records," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, p. 149, Dec. 2022, doi: 10.1186/s12911-022-01894-7.

[36] V. Buhrmester, D. Münch, and M. Arens, "Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey," *Mach. Learn. Knowl. Extr.*, vol. 3, no. 4, pp. 966–989, Dec. 2021, doi: 10.3390/make3040048.

[37] A. Saibene, M. Assale, and M. Giltri, "Expert systems: Definitions, advantages and issues in medical field applications," *Expert Syst. Appl.*, vol. 177, p. 114900, Sep. 2021, doi: 10.1016/j.eswa.2021.114900.

[38] A. Alharan, N. Ali, and Z. Algelal, "An Adaptive Neuro-Fuzzy Inference System And Principal Component Analysis: A Hybridized Method For Viral Hepatitis Diagnosis System," Apr. 2022, doi: 10.24412/2520-6990-2022-19142-4-10.

[39] C. W. McMonnies, "Why the symptoms and objective signs of dry eye disease may not correlate," *J. Optom.*, vol. 14, no. 1, pp. 3–10, Jan. 2021, doi: 10.1016/j.optom.2020.10.002.

[40] D. A. Dharmawan, "Assessing fairness in performance evaluation of publicly available retinal blood vessel segmentation algorithms," *J. Med. Eng. Technol.*, vol. 45, no. 5, pp. 351–360, Jul. 2021, doi: 10.1080/03091902.2021.1906342.

[41] K. Khalil, O. Eldash, A. Kumar, and M. Bayoumi, "Designing Novel AAD Pooling in Hardware for a Convolutional Neural Network Accelerator," *IEEE Trans Very Large Scale Integr VLSI Syst.*, vol. 30, no. 3, pp. 303–314, Mar. 2022, doi: 10.1109/TVLSI.2021.3139904.

[42] A. A. Abdulsahib, M. A. Mahmoud, M. A. Mohammed, H. H. Rasheed, S. A. Mostafa, and M. S. Maashi, "Comprehensive review of retinal blood vessel segmentation and classification techniques: intelligent solutions for green computing in medical images, current challenges, open issues, and knowledge gaps in fundus medical images," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 10, no. 1, p. 20, Dec. 2021, doi: 10.1007/s13721-021-00294-7.

[43] A. G. Figueroa *et al.*, "Levodopa Positively Affects Neovascular Age-Related Macular Degeneration," *Am. J. Med.*, vol. 134, no. 1, pp. 122-128.e3, Jan. 2021, doi: 10.1016/j.amjmed.2020.05.038.

[44] P. Vivek, S. Goel, R. Nijhawan, and S. Gupta, "CNN Models and Machine Learning Classifiers for Analysis of Goiter Disease," in *2022 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1–6, Feb. 2022, doi: 10.1109/SCEECS54111.2022.9740737.

[45] F. lahmood Hameed and O. Dakkak, "Brain Tumor Detection and Classification Using Convolutional Neural Network (CNN)," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–7, Jun. 2022, doi: 10.1109/HORA55278.2022.9800032.

[46] A. Samanta, A. Saha, S. C. Satapathy, S. L. Fernandes, and Y.-D. Zhang, "Automated detection of diabetic retinopathy using convolutional neural networks on a small dataset," *Pattern Recognit. Lett.*, vol. 135, pp. 293–298, Jul. 2020, doi: 10.1016/j.patrec.2020.04.026.

[47] A. Mumuni and F. Mumuni, "CNN Architectures for Geometric Transformation-Invariant Feature Representation in Computer Vision: A Review," *SN Comput. Sci.*, vol. 2, no. 5, p. 340, Sep. 2021, doi: 10.1007/s42979-021-00735-0.

[48] M. Y. Demirci, N. Beşli, and A. Gümüşçü, "Efficient deep feature extraction and classification for identifying defective photovoltaic module cells in Electroluminescence images," *Expert Syst. Appl.*, vol. 175, p. 114810, Aug. 2021, doi: 10.1016/j.eswa.2021.114810.

[49] T. Rezaie *et al.*, "Adult-Onset Primary Open-Angle Glaucoma Caused by Mutations in Optineurin," *Science (1979)*, vol. 295, no. 5557, pp. 1077–1079, Feb. 2002, doi: 10.1126/science.1066901.

[50] G. Boraiah and A. Chhabra, "Conventional and Advanced Imaging Evaluation of Spine," in *Multidisciplinary Spine Care*, 73–107, 2022, doi: 10.1007/978-3-031-04990-3_4.

[51] C. Zhang *et al.*, "Acceleration of three-dimensional diffusion magnetic resonance imaging using a kernel low-rank compressed sensing method," *Neuroimage*, vol. 210, p. 116584, Apr. 2020, doi: 10.1016/j.neuroimage.2020.116584.

[52] C. G. Schwarz *et al.*, "Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives," *Neuroimage*, vol. 231, p. 117845, May 2021, doi: 10.1016/j.neuroimage.2021.117845.

[53] M. Heinsch *et al.*, "Supporting friends and family of adults with a primary brain tumour: A systematic review," *Health Soc Care Community*, vol. 30, no. 3, pp. 869–887, May 2022, doi: 10.1111/hsc.13586.

[54] Z. Huang, Y. Zhao, Y. Liu, and G. Song, "AMF-Net: An adaptive multisequence fusing neural network for multi-modality brain tumor diagnosis," *Biomed. Signal Process Control*, vol. 72, p. 103359, Feb. 2022, doi: 10.1016/j.bspc.2021.103359.

[55] B. M. Pacheco, G. de S. e Cassia, and D. Silva, "Towards fully automated deep-learning-based brain tumor segmentation: Is brain extraction still necessary?," *Biomed. Signal Process Control*, vol. 82, p. 104514, Apr. 2023, doi: 10.1016/j.bspc.2022.104514.

[56] R. Vankdothu, M. A. Hameed, and H. Fatima, "A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method," *Computers and Electrical Engineering*, vol. 101, p. 107960, Jul. 2022, doi: 10.1016/j.compeleceng.2022.107960.

[57] S. K. Rajeev, M. Pallikonda Rajasekaran, G. Vishnuvarthanan, and T. Arunprasath, "A biologically-inspired hybrid deep learning approach for brain tumor classification from magnetic resonance imaging using improved gabor wavelet transform and Elmann-BiLSTM network," *Biomed. Signal Process Control*, vol. 78, p. 103949, Sep. 2022, doi: 10.1016/j.bspc.2022.103949.

[58] R. Sindhiya Devi, B. Perumal, and M. Pallikonda Rajasekaran, "A hybrid deep learning based brain tumor classification and segmentation by stationary wavelet packet transform and adaptive kernel fuzzy c means clustering," *Advances in Engineering Software*, vol. 170, p. 103146, Aug. 2022, doi: 10.1016/j.advengsoft.2022.103146.

[59] X. Lei, X. Yu, J. Chi, Y. Wang, J. Zhang, and C. Wu, "Brain tumor segmentation in MR images using a sparse constrained level set algorithm," *Expert Syst. Appl.*, vol. 168, p. 114262, Apr. 2021, doi: 10.1016/j.eswa.2020.114262.

[60] K. V. S. Et. al., "Study and Analysis of Various Automatic Brain Tumour Segmentation andClassification: A Challenging Overview," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 10, pp. 6106–6115, May 2021, doi: 10.17762/turcomat.v12i10.5447.

[61] V. D. Ugale, S. S. Pawar, and S. Pawar, "Brain Tumour Detection using Image Processing," in *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 667–672, Apr. 2022, doi: 10.1109/CSNT54456.2022.9787643.

[62] N. M. Ghadi and N. H. Salman, "Deep Learning-Based Segmentation and Classification Techniques for Brain Tumor MRI: A Review," *Journal of Engineering*, vol. 28, no. 12, pp. 93–112, Dec. 2022, doi: 10.31026/j.eng.2022.12.07.

[63] H. Mzoughi, I. Njeh, M. Ben Slima, A. Ben Hamida, C. Mhiri, and K. Ben Mahfoudh, "Towards a computer aided diagnosis (CAD) for brain MRI glioblastomas tumor exploration based on a deep convolutional neuronal networks (D-CNN) architectures," *Multimed Tools Appl.*, vol. 80, no. 1, pp. 899–919, Jan. 2021, doi: 10.1007/s11042-020-09786-6.

[64] F. Gao *et al.*, "AD-NET: Age-adjust neural network for improved MCI to AD conversion prediction," *Neuroimage Clin.*, vol. 27, p. 102290, 2020, doi: 10.1016/j.nicl.2020.102290.