

---

# AN ARTIFICIAL INTELLIGENCE FRAMEWORK FOR MEASURING HUMAN SPINE AGING USING MRI

---

Roozbeh Bazargani<sup>1</sup> Saqib Abdullah Basar<sup>1</sup> Daniel Daly-Grafstein<sup>1</sup> Rodrigo Solis Pompa<sup>1</sup> Soojin Lee<sup>1</sup>  
 Saurabh Garg<sup>1</sup> Yuntong Ma<sup>1</sup> John A. Carrino<sup>2,3</sup> Siavash Khallaghi<sup>1,\*</sup> Sam Hashemi<sup>1,\*</sup>

<sup>1</sup>*Prenuvo, Redwood City, CA, USA*

<sup>2</sup>*Weill Cornell Medical College, New York, NY, USA*

<sup>3</sup>*Hospital for Special Surgery, New York, NY, USA*

<sup>\*</sup>*Co-senior authors*

## ABSTRACT

The human spine is a complex structure composed of 33 vertebrae. It holds the body and is important for leading a healthy life. The spine is vulnerable to age-related degenerations that can be identified through magnetic resonance imaging (MRI). In this paper we propose a novel computer-vision-based deep learning method to estimate spine age using images from over 18,000 MRI series. Data are restricted to subjects with only age-related spine degeneration. Eligibility criteria are created by identifying common age-based clusters of degenerative spine conditions using uniform manifold approximation and projection (UMAP) and hierarchical density-based spatial clustering of applications with noise (HDBSCAN). Model selection is determined using a detailed ablation study on data size, loss, and the effect of different spine regions. We evaluate the clinical utility of our model by calculating the difference between actual spine age and model-predicted age, the spine age gap (SAG), and examining the association between these differences and spine degenerative conditions and lifestyle factors. We find that SAG is associated with conditions including disc bulges, disc osteophytes, spinal stenosis, and fractures, as well as lifestyle factors like smoking and physically demanding work, and thus may be a useful biomarker for measuring overall spine health.

## 1 Introduction

The spine is a complex anatomical structure consisting of numerous vertebrae (nominally, 7 cervical, 12 thoracic, and 5 lumbar) and motion segments that serves as both a supportive axis and a protective encasement for the spinal cord. It is integral to biomechanical function and core stability, enabling a diverse range of movements, and plays a fundamental role in facilitating the transmission of neural signals by distributing the peripheral nervous system. Age-related changes and pathological alterations can substantially disrupt these functions, leading to reduced mobility, sensory impairments, loss of motor function, and decreased quality of life. The aging process of the spine is multifactorial, driven by an interplay of genetic predispositions, underlying pathologies, and lifestyle factors [1]. Radiological imaging is crucial for the identification of spinal conditions, which may involve various components, including intervertebral discs, facet (zygapophyseal) joints, neural canals, muscles, tendons, ligaments, and bony structures. Magnetic resonance imaging (MRI) provides superior visualization of the musculoskeletal system, enabling the assessment and quantification of age-related morphological changes. Consequently, there is an increasing interest in the development of automated methods for spine age estimation utilizing imaging to determine what a healthy spine typically looks like at different stages of life. The objective of this study was to explore the feasibility of predicting spine age using deep learning models applied to whole-spine MRI images.

### 1.1 Spine Maturation and Degeneration

Spine degeneration and stratification of spine patients have been of interest to radiologists. Kim et al. [2] introduced the classification of lumbar disc herniation from MR images in 1992. Pfirrmann et al. [3] developed a classification system for lumbar intervertebral disc degeneration based on MR images. They showed a T2-weighted MRI is reliable for grading disc degeneration. Riesenburger et al. [4] proposed a novel classification method for lumbar disc degeneration

that incorporates endplate changes, the presence of a high-intensity zone, and a reduction in disc height. Gille et al. [5, 6] proposed a classification system for degenerative spondylolisthesis of the lumbar spine that was consistent with age and health-related quality of life scales.

## 1.2 Previous Work in Spine Degeneration Detection

Previous work shows the feasibility of determining spine degeneration from MR images using deep learning. Lu et al. [7] proposed a deep learning method to determine the severity of spinal canal stenosis, left foraminal stenosis, and right foraminal stenosis. The method consisted of 1) segmentation and labeling of vertebrae using U-Net; 2) extracting disc-level image volumes from sagittal and axial views; and 3) passing them through a Convolutional Neural Network (CNN) to determine the severity. SpineOne [8] is a one-stage framework that detects lumbar vertebrae and discs and classifies them into normal and degenerative. Hallinan et al. [9] proposed a deep learning method to automate the detection and grading of the central canal, lateral recess, and neural foraminal stenosis in lumbar spine MR images. They used a combination of faster Region-based Convolutional Neural Network (RCNN) to detect the region of interest and a CNN classifier to grade the condition. Zheng et al. [10] quantitated lumbar disc degeneration from MRI based on the feature extracted after segmenting vertebrae, discs, and the spinal cord and extracting feature points. Yi et al. [11] used a combination of CNN and transformer to detect degenerative diseases in the cervical and lumbar regions from T2-weighted MR images taken in sagittal and axial views. Chen et al. [12] proposed a two-stage framework using Mask RCNN. The disc and vertebrae were localized based on sagittal view and the degenerative conditions were classified based on sagittal and axial views.

These studies suggest that spine degeneration can be detected using deep learning approaches that mimic a visual inspection of the spine. Given that these conditions are correlated with aging, we hypothesize that whole spine MRI contains information that can be leveraged by deep learning models to predict spine age.

## 1.3 Previous Work in Age Estimation

In the field of computer vision, most age estimation methods focused on two-dimensional images of the face [13, 14, 15, 16]. In medical imaging estimation of age based on brain MRI scans has been extensively researched [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. The estimation can be achieved through two approaches [17]:

- Regression method: The model predicts a continuous value representing the chronological age.
- Bins method: The model uses several bins and predicts the bin containing the chronological age, outputting the probability of the brain age belonging to each bin. The final age is computed as the expected value.

Various models and loss functions have been proposed to tackle this task. Ordinal distance encoded regularization loss has been added to the cross entropy loss to train a CNN [19]. Attention-guided deep learning was used to predict the gestational age from T2-weighted MRI [20]. The Extreme Gradient Boosting (XGBoost) regression algorithm was applied to approximately 42,000 T1-weighted brain MRI series [23]. Peng et al. [24] used a lightweight fully convolutional structure based on VGGNet [29] architecture in T1-weighted structural MRI (sMRI). Armanious et al. [25] proposed Age-Net, a deep CNN with a novel iterative data-cleaning algorithm to separate atypical-aging patients in T1-weighted MRI. Two-Stage-Age-Network was proposed to improve the brain age estimation in T1-weighted MRI [26]. The first-stage network estimates the brain age based on T1-weighted MRI and sex. Then, the age estimate is refined in the second-stage network. The inputs to the second stage are MRI image, sex, and the discretized brain age estimated by the first-stage network and the output is the residual that is added to the discretized estimated brain age. Ensemble learning was applied to sMRI to improve the performance [27]. Three separate ensemble frameworks were used for young, middle-aged, and older age groups to mitigate the dependency of model accuracy on age groups. Each framework contained four models: a support vector machine, a CNN, GoogleNet [30], and ResNet [31]. In summary, these studies demonstrate the feasibility of assigning organ-specific biological age using deep learning methods on MR images.

In the context of the spine, there has been research on the relationship between spine conditions, age, and occupation [32, 33, 34]. Some studies [35, 36] utilized machine learning techniques to estimate spine age from MR images. However, the performance was low with  $R^2 = 0.28$  and a mean absolute error of 10.28 years for spine age compared to chronological age in Sneath et al. study [36]. Moreover, these studies were limited by small datasets (60 and 70 MRI scans for training with test sets of 9 and 10 MRI scans, respectively) and relied on feature engineering rather than directly using the MRI scans as inputs. Furthermore, none of these studies utilized a deep learning approach and instead relied on classical methods such as random forests, extreme gradient boosting trees, and support vector machines.

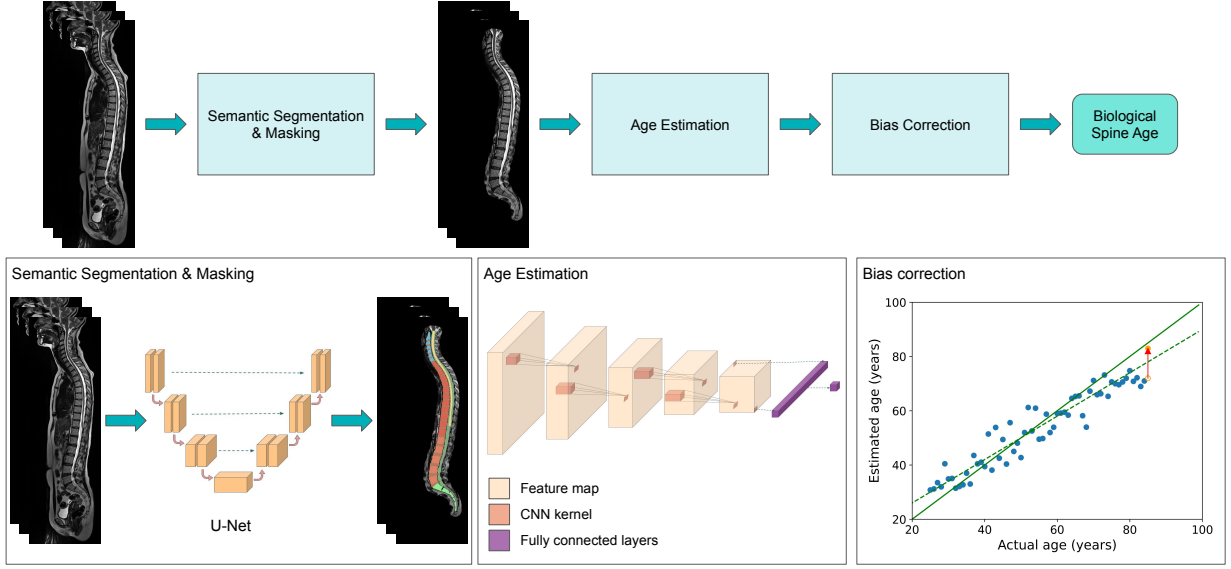


Figure 1: Overview of the three steps of the model to produce biological spine age. In the first step, semantic segmentation of the spine is generated and masked using a nnUnet model [37]. Next, the masked spine series are passed to a DCNN model consisting of multiple blocks of 3D convolution, batch-normalization, max-pooling layers, and a linear layer for the final prediction. The details of the model and layers are available in Table 2. In the last step, we perform bias correction to produce bias-corrected spine age.

## 1.4 Contributions

We propose a Deep Convolutional Neural Network (DCNN) for predicting spine age based on a large dataset of sagittal T2-weighted MRI series.

- To the best of our knowledge, this is the first work that utilizes MRI as input to a deep learning model for spine age estimation, achieving an  $R^2$  performance of 0.87 compared to 0.28 in prior studies. This demonstrates that spine age can be estimated from T2-weighted MRI with high accuracy.
- We utilized an extensive dataset comprising of more than 18,000 series from more than 17,000 participants with over 8,500 test samples of normal and abnormal MR images. This significantly surpasses previous work in spine age estimation, which used only 70 series in total and 10 series in their test set.
- Our study demonstrates that the difference between chronological age and spine age, hereafter referred to as Spine Age Gap (SAG), is related to spinal conditions, pathologies, and lifestyle factors. This finding suggests that the SAG can serve as a valuable metric for assessing spine health.

## 2 Material and Methods

### 2.1 Overview

Figure 1 shows an overview of our proposed spine age estimation model. The pipeline consists of three steps: 1) semantic segmentation and masking; 2) age estimation; and 3) bias correction. Next, we discuss the data and each of these steps in detail.

### 2.2 Data

We utilized a comprehensive dataset comprising of 18,070 3D T2-weighted whole spine MRI series. This dataset was acquired from 17,394 individuals with the sagittal view as the imaging plane. The scans were collected over a thirteen-year period, from 2011 to 2024, using 19 Philips and Siemens machines across 10 clinics in North America. The dataset included individuals aged 25 to 84 years, including those with supernumerary vertebrae (anomalous enumeration

Table 1: Distribution of males and females in normal and abnormal series across age groups.

Age bracket (years)	No. normal series		No. abnormal series	
	male	female	male	female
30 (25-34)	561	521	404	411
40 (35-44)	1306	1238	1249	1252
50 (45-54)	1617	1379	857	949
60 (55-64)	1264	1299	645	683
70 (65-74)	598	622	377	378
80 (75-84)	122	84	132	122
Total	5468	5143	3664	3795
	10611		7459	

of vertebrae). Individuals were divided into six age groups, each spanning a 10-year interval, as shown in Table 1. The classification for 'normal' and 'abnormal,' as referenced in Table 1, are explained in detail in the following section.

### 2.3 Establishing Normal Spines Based on Reports

To train a DCNN model for predicting chronological age from MRI input, the dataset must be restricted to subjects who exhibit only age-related spine degeneration. To this end, we propose data-driven eligibility criteria to define participants who have a normal spine with respect to their age.

We follow a clustering approach for identifying normal participants based on spine conditions in their radiology report. This includes spinal structural and canal pathologies, as well as degenerative conditions for each vertebra.

The seven spinal structural and canal pathologies are: 1) bone lesion; 2) congenital spinal canal narrowing; 3) cord abnormalities; 4) fracture; 5) soft tissue edema; 6) spinal stenosis, and 7) spondylolisthesis.

The eight degenerative spinal conditions consist of: 1) disc bulge; 2) disc osteophyte complex; 3) uncovertebral osteophyte; 4) protrusion; 5) extrusion; 6) desiccation; 7) endplate change; and 8) annular fissure. In total, we looked at six cervical (C2-C7), thirteen thoracic (T1-T12, including T13 for cases with supernumerary vertebrae), and seven lumbar (L1-L5, including L6-L7 for cases with supernumerary vertebrae) vertebrae that are  $6 + 13 + 7 = 26$  vertebrae in total. Thus, we had  $26 \times 8 + 7 = 215$  features for each report associated with each series.

This 215 feature vector is extremely sparse for the typical participant. Furthermore, radiologists frequently perform regional assessment of the spine and only report the most pertinent pathology in each region. We used this heuristic to convert this sparse feature vector into a dense representation. To this end, we aggregated affected vertebrae by region and severity of conditions. This reduced the number of features into  $3 \times 8 \times 4 = 96$  for three regions (cervical, thoracic and lumbar) for eight degenerative spinal conditions and four severity values (mild, moderate, severe, and near complete). However, not all conditions are associated with all severity values listed above, which results in a smaller aggregated feature vector of size 60 for spinal conditions. By adding spinal structural and canal pathologies conditions we arrive at a final vector of size 67.

In order to visualize this feature vector we performed a Uniform Manifold Approximation and Projection (UMAP) [38] dimensionality reduction to planar features for each age bracket. However, for our feature vector the default Euclidean distance used in UMAP evaluates different conditions and the frequency of conditions equally. In order to address this issue, we used the Canberra distance as defined below:

$$d(p, q) = \sum_{i=1}^m \frac{|p_i - q_i|}{|p_i| + |q_i|}, \quad (1)$$

where  $m$  is the total number of conditions,  $p$  and  $q$  are vectors of the conditions. This formulation ensures that the distance between two cases where a specific condition has one more count is less than having a different condition. For instance, if participant 1 ( $p_1$ ) has 3 mild disc bulges, participant 2 ( $p_2$ ) has 4 mild disc bulges, and participant 3 ( $p_3$ ) has 3 mild disc bulges and 1 fracture,  $d(p_1, p_2) = \frac{1}{7}$  and  $d(p_1, p_3) = 1$ .

Next, we applied Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [39] to the reduced dimensions by UMAP to cluster samples. Within each age group clusters with more than 15% of the population were considered normal, and clusters with less than 15% were considered abnormal.

Table 2: Layers of the model including the shape and number of parameters (weights + biases). Total number of parameters is 2,950,401.

	Layer	Shape	Num Parameters
block1	conv3d	[32, 1, 3, 3, 3]	$864 + 32 = 896$
	batchnorm3d	[32]	$32 + 32 = 64$
	relu		
	maxpool3d		
block2	conv3d	[64, 32, 3, 3, 3]	$55296 + 64 = 55360$
	batchnorm3d	[64]	$64 + 64 = 128$
	relu		
	maxpool3d		
block3	conv3d	[128, 64, 3, 3, 3]	$221184 + 128 = 221312$
	batchnorm3d	[128]	$128 + 128 = 256$
	relu		
	maxpool3d		
block4	conv3d	[256, 128, 3, 3, 3]	$884736 + 256 = 884992$
	batchnorm3d	[256]	$256 + 256 = 512$
	relu		
	maxpool3d		
block5	conv3d	[256, 256, 3, 3, 3]	$1769472 + 256 = 1769728$
	batchnorm3d	[256]	$256 + 256 = 512$
	relu		
	maxpool3d		
top	conv3d	[64, 256, 1, 1, 1]	$16384 + 64 = 16448$
	batchnorm3d	[64]	$64 + 64 = 128$
	relu		
	maxpool3d		
prediction	linear	[1, 64]	$64 + 1 = 65$

## 2.4 Spine Age Estimation with Deep Convolutional Neural Network

The field of view in a whole spine MRI spans multiple organs that age at different rates. In order to disentangle the spine from the rest of the organs, we first segment the spine using a semantic segmentation model [37]. This process generates a segmentation mask that encompasses the cervical, thoracic, lumbar, and sacral vertebrae, intervertebral discs, ribs, cerebrospinal fluid, and the spinal cord. This mask is dilated and used to remove other regions from the MR image. In order to decrease the spatial variability of samples in our dataset, we resample all series to a common spacing of  $0.9 \times 0.9 \times 3 \text{ mm}^3$ . Subsequently, we center cropped / padded all images to a fixed size of  $384 \times 793 \times 14$ .

Our DCNN model was inspired by previous work in brain age estimation [40, 41]. The model is shown in Table 2. Mean Squared Error (MSE) loss was used to train the model.

After training, we used the validation data to correct the bias using Cole’s method [18] as used in previous studies [24, 17]. We computed the slope  $\alpha$  and intercept  $\beta$  of the fitted line to the samples using linear regression

$$\hat{Y} = \alpha Y + \beta, \quad (2)$$

where  $Y$  and  $\hat{Y}$  represent chronological and predicted age, respectively. The bias-corrected predicted age  $\hat{Y}_c$  is computed as

$$\hat{Y}_c = \frac{\hat{Y} - \beta}{\alpha}. \quad (3)$$

## 3 Experimental Settings

### 3.1 Normal Data

We only used the images from individuals clustered as normal (see Section 4.1) within their age group for training our DCNN model. This resulted in 10,611 series for the development of the spine age estimation model. We divided the dataset into training, validation, and test sets with 8,491, 1,051, and 1,069 samples, respectively. The samples were split while maintaining the gender and age ratios, as shown in Figure 2. To evaluate the relationship of spine

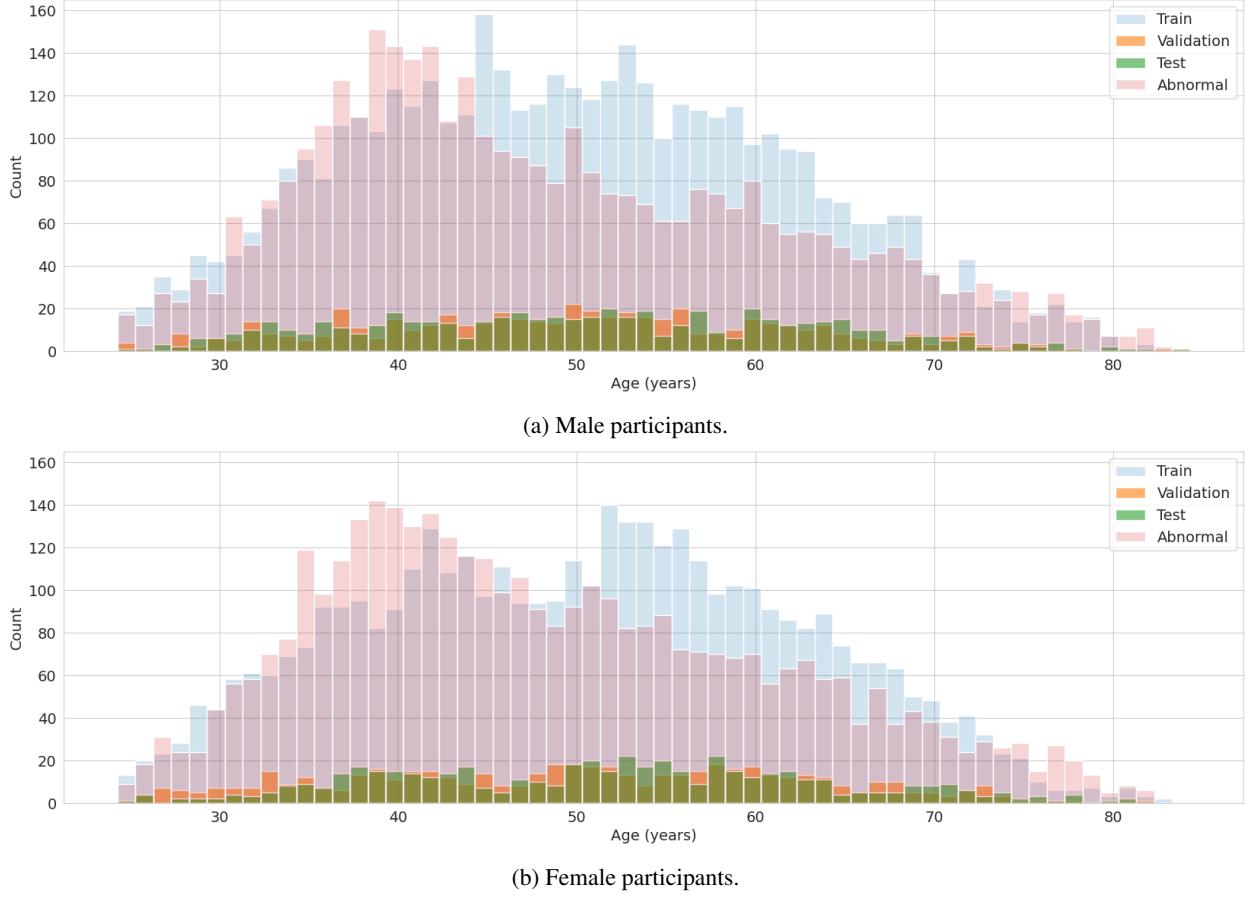


Figure 2: Train, validation, and test set histogram based on gender across age brackets.

conditions and lifestyle with predicted spine age, we mixed the abnormal data with the test set, resulting in a set of  $7459 + 1069 = 8528$  samples for the full-test set.

### 3.2 Sample size, Loss, and Spine regions

To assess the proposed model, we compared it with models trained on a smaller number of samples, different loss, and different regions of the spine. In particular, we trained the same model on 85 (85 train and 10 validation) and 850 (850 train and 100 validation) samples. In addition, we trained the model using smooth-L1 loss. Similar to L1 loss, smooth-L1 loss helps mitigate the sensitivity to noisy series. We also trained the model on masked series that only included cervical, thoracic, and lumbar regions. Having segmentations of cervical, thoracic, and lumbar vertebrae, we masked those regions after dilation and trained models on the masked series.

The metrics used for comparison are mean absolute error (MAE), age bracket weighted MAE (WMAE), and  $R^2$  between the actual and predicted age. A lower absolute error indicates better performance, as it shows that the model is predicting age more accurately. MAE is computed as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (4)$$

where  $y_i$  and  $\hat{y}_i$  are  $i$ -th participant's chronological and predicted age and  $n$  is the number of participants.

WMAE is calculating by computing the MAE in each age category, then taking the global mean of these age category errors as:

$$\text{WMAE} = \frac{1}{N} \sum_{j=1}^N \left( \frac{1}{n_j} \sum_{i=1}^{n_j} |y_i - \hat{y}_i| \right), \quad (5)$$

where  $N$  is the number of age brackets and  $n_j$  is the number of samples in each age bracket. This allows all age categories to contribute equally in model evaluation regardless of the number of samples.

$R^2$  has a range of  $-\infty$  to 1 and the higher values show better performance. It is computed as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (6)$$

where  $\bar{y}$  is the mean of participants' age.

### 3.3 Hyperparameters

#### 3.3.1 UMAP

We set the number of neighbors to fifteen, and the minimum distance to zero.

#### 3.3.2 HDBSCAN

We set the minimum number of samples in a cluster to 1% of the population in each age bracket. If a cluster has a smaller number of samples, it is considered an outlier. We set the number of samples in a neighborhood for a sample to be considered as a core sample to 5. Finally, we set the distance threshold that merges the clusters whose distance is below this threshold to 0.3 for the 70 and 80 age brackets, 0.7 for 40 and 60, and 1 for 30 and 50 based on the distribution of the sample points in the UMAP plot.

#### 3.3.3 DCNN

For training the model, the batch size was set to two. Adam's optimizer was used with a learning rate of 0.01. We also used a reduce learning rate on plateau scheduler with a factor of 0.3 and a patience of five.

## 4 Results and Discussion

### 4.1 Establishing the Normal Spine

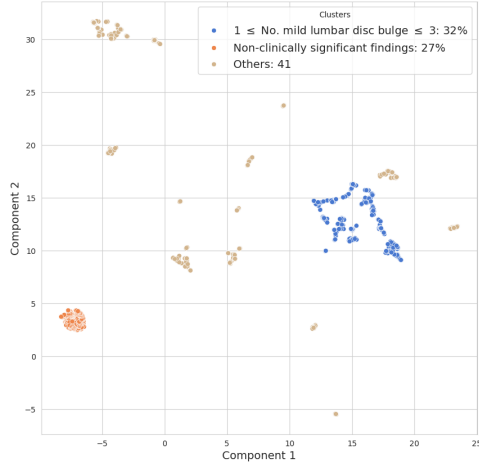
Our 15% population threshold for age-based clusters described in Section 2.3 resulted in 32% to 54% of spines defined as abnormal across age brackets, as shown in Figure 3. We labeled normal clusters based on the primary conditions that were observed in the majority of samples in the cluster.

In the 30-year-old age bracket, having one to three mild lumbar disc bulges is the dominant cluster (32% of the population). The next cluster represents 27% of the population where most participants did not have a clinically significant finding. In the 40-year-old population, the population percentage of mild lumbar disc bulges is increased but divided into two groups based on having less or equal to 3 mild cervical disc osteophytes or between 1 to 3 mild cervical disc bulges. Moreover, the percentage of the population without a clinically significant finding reduced from 27% to 15%.

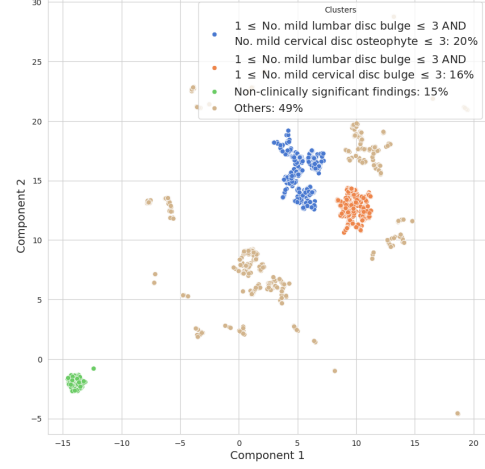
The two clusters based on mild lumbar disc bulges either with mild cervical disc osteophyte or mild cervical disc bulge exist in the 50-, 60-, and 70-year-old brackets as well. The only difference is the number of occurrences in normal spines increases with age. Furthermore, the non-clinically significant findings cluster is not considered normal in these age groups. Finally, in the 80-year-old group, mild cervical uncovertebral osteophyte is another condition in addition to mild lumbar disc bulge and mild cervical disc osteophyte.

### 4.2 Model Results

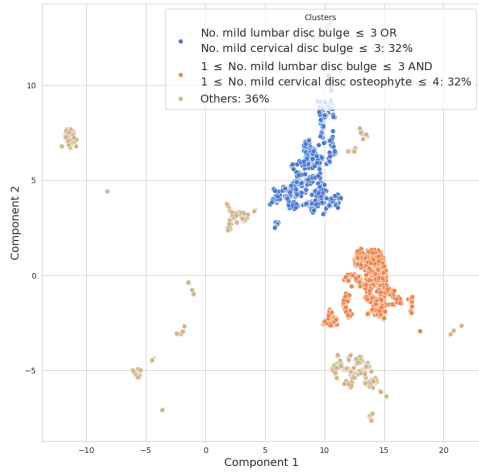
Model results comparing different age estimation models are shown in Table 3. These results are based on model evaluation of the 1069 normal spine series in the test set. While looking at the performance of the models with and without bias correction in Table 3, we note that although bias correction has increased the MAE, it has reduced the WMAE for all models except the proposed model where the WMAE remained unchanged. We see this because our data has an imbalanced age distribution with many cases close to the overall mean age in our sample (Figure 2). Uncorrected predictions tend to be biased towards these mean ages, while bias correction removes the bias toward the mean and makes the error uniform. As a result bias correction increased the MAE, but the WMAE that computes error uniformly



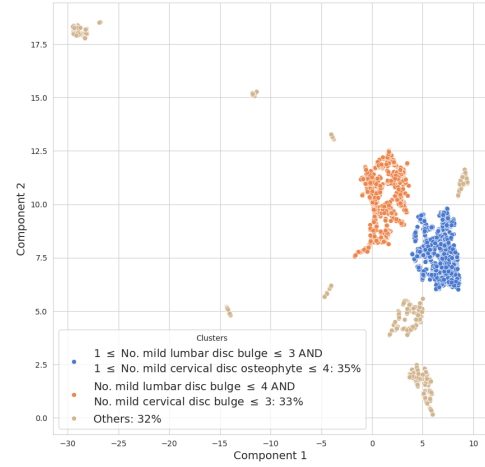
(a) 30-year-old age bracket



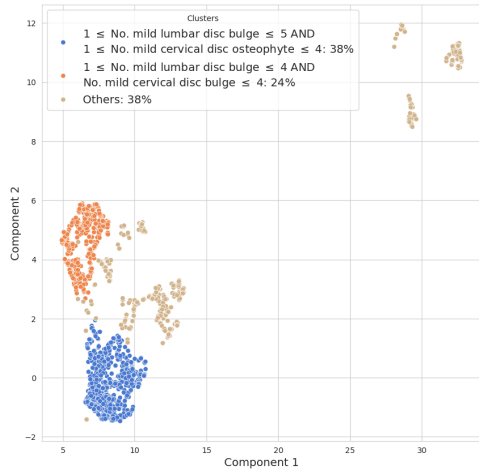
(b) 40-year-old age bracket



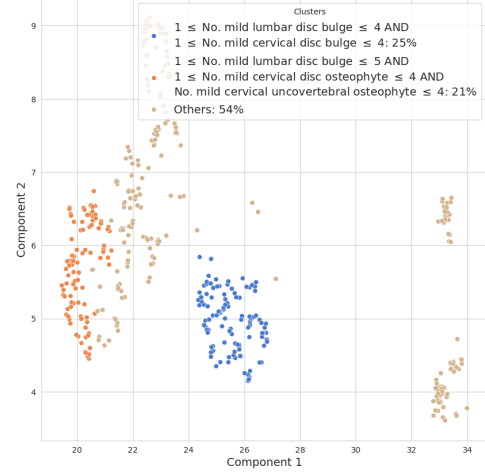
(c) 50-year-old age bracket



(d) 60-year-old age bracket



(e) 70-year-old age bracket



(f) 80-year-old age bracket

Figure 3: Clusters based on UMAP-reduced spine conditions. Only clusters that were more than 15% of the population were kept and the rest merged into one. We labeled the clusters based on the dominant conditions in the cluster. The population percentage of each cluster is written in front of the label.



Table 3: Comparison of age estimation models on normal data based on Mean Absolute Error (MAE) in years,  $R^2$ , and age bracket weighted MAE (WMAE) with and without Bias Correction (BC). Bold numbers represent the best results, while underlined numbers represent second-best results.

Model		Without BC			With BC		
		MAE	$R^2$	WMAE	MAE	$R^2$	WMAE
Data size	85 samples	9.77	-0.03	11.51	9.76	-0.04	10.55
	850 samples	4.82	0.75	5.57	5.33	0.69	5.50
Loss	Smooth-L1	<u>3.60</u>	<u>0.86</u>	<u>4.47</u>	<u>3.94</u>	<u>0.83</u>	<u>4.00</u>
	Cervical	4.93	0.75	5.68	5.57	0.65	5.61
Regions	Thoracic	4.02	0.82	4.77	4.57	0.77	4.58
	Lumbar	3.89	0.83	4.67	4.35	0.79	4.57
Proposed model		<b>3.47</b>	<b>0.87</b>	<b>3.60</b>	<b>3.67</b>	<b>0.85</b>	<b>3.60</b>

Table 4: Average years between scans, SAG intraclass correlation coefficient (ICC) and 95% bootstrap confidence intervals of patients with two scans in our cohort.

Subgroup	n	Avg Years	ICC
Full test set	303	1.59	0.73 (0.68, 0.78)
Male	172	1.60	0.72 (0.64, 0.79)
Female	131	1.56	0.74 (0.63, 0.81)

and independent of the imbalanced distribution across age brackets, declined. This result suggests that performing bias correction improved age estimation for our DCNN model. For the remainder of this manuscript, the spine age estimation refers to bias corrected values.

#### 4.2.1 Size of Dataset

To evaluate the effect of dataset size on model performance, we tested our model on subsets of 85 and 850 samples in addition to the full dataset of 8491 series used for training the proposed model. The 85 samples subset represents 1% of the data and is larger than the largest dataset previously used for predicting spine age [35, 36]. The 850-sample model represents the performance using 10% of the data. It can be observed in Table 3 that increasing the training set size improves performance. Training with 85, 850, and 8491 series resulted in  $R^2$  of  $-0.04$ ,  $0.69$ , and  $0.85$ , respectively.

#### 4.2.2 Choice of the Loss Function

We trained the proposed DCNN model using the MSE (L2) loss vs. smooth-L1 loss and compared their respective  $R^2$  on the test set. The results of this experiment are shown in Table 3. L2 loss showed a  $0.02$  improvement in  $R^2$  and a reduction of  $0.27$  and  $0.40$  years in MAE and WMAE, respectively when compared to smooth-L1 loss.

#### 4.2.3 Region Specific vs. Whole Spine Model

*Are specific regions of the spine a better indicator of spine age?* In order to answer this question, we trained our model on masked inputs from different regions and compared their performance against the proposed whole spine model. The segmentation model [37] is capable of detecting lumbar, thoracic, and cervical vertebrae. We used this feature to keep only one of these regions and mask the rest of the image. As seen in Table 3, the lumbar region exhibits improved performance compared to cervical and thoracic. This shows that aging is more apparent in the lumbar back region [42]. Furthermore, the model trained on the whole spine region shows an improved performance of  $0.06$  in  $R^2$  compared to the model trained only on the lumbar region. This suggests that all regions of the spine are important for the assessment of the biological spine age.

#### 4.2.4 Repeat Scan Stability

We evaluated model predictions on the 303 individuals who received two scans in the 8528 series in our test set. We expect repeat scans done on the same individual in a short time frame to have similar predicted ages. Results of these multiple scan predictions are shown in Table 4. We see moderately strong stability in our spine age predictions between scans, with an intraclass correlation of  $0.73$ . We note that we do not expect perfect stability given there was an average of  $1.6$  years between scans and we expect biological and chronological aging differences between scans.

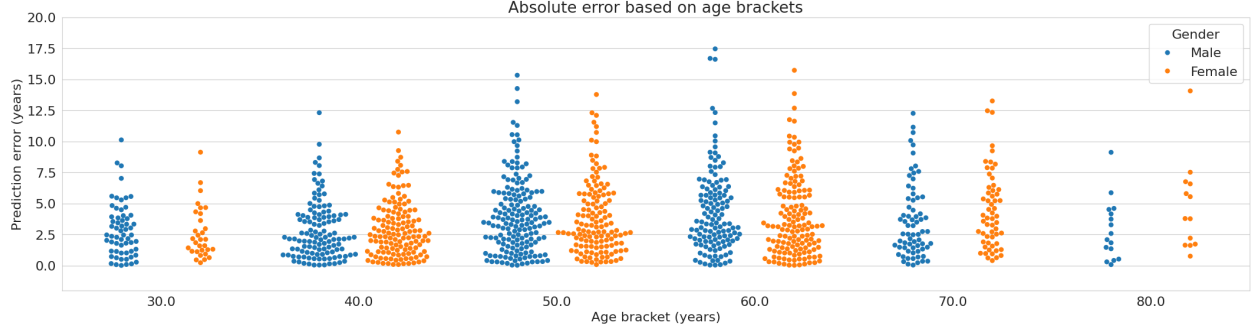


Figure 4: Absolute error on the normal test set grouped based on gender and age bracket

### 4.3 Analysis of Large Spine Age Differences

In Figure 4, we illustrate the absolute error of the normal test set in years based on gender and age brackets. Upon conducting inference on the full-test set, we identified 30 instances where the discrepancy between chronological age and spine age exceeded 15 years. A thorough examination of these cases, supplemented by radiology reports and expert feedback, revealed that 23 of those cases were consistent with expected clinical patterns. For instance, we had a healthy 77-year-old participant whose spine age was 37. Of the remaining 7 samples, 2 of them had artifacts in the MR images, the segmentation model failed in 1 of the instances to properly mask the spine, and the spine age estimation model failed in 4 cases. The failure description is as follows:

1. A 77-year-old participant whose spine age was estimated to be 39 years. However, based on expert feedback a spine age estimation between 45 to 50 years appears to be more consistent with the participant’s spine condition. There are areas of increased intensity, such as C6 vertebra. Looking at the Grad-CAM the model has a high attention on C6 but was unable to predict the spine age more accurately.
2. A 80-year-old participant with a vertebral fracture whose spine age was estimated at 63 years. The model has mostly focused on the disc bulges and has failed to attend to the vertebral fracture.
3. A 57-year-old participant whose spine age was estimated at 74 years. There is a loss of thoracic and cervical curvature to some degree but there is nothing specific to explain the age overestimation.
4. A 61-year-old participant whose spine age was estimated at 77 years. There is loss of intervertebral disc space at upper thoracic levels and slight rightward scoliosis but nothing special to indicate the overestimated spine age. After looking into Grad-CAM heatmaps, we realized that the model has focused on regions not specific to the spine, such as areas near the brain.

### 4.4 Grad-CAM Results

To interpret the model predictions, we used Grad-CAM [43] to assign heatmaps to each region based on the contribution of the region in the final output. We generated two-dimensional heatmaps based on the fifth block of the model. The heatmaps are adjusted for a better contrast based on  $f(x) = \max(\ln(288x), 1)$ . The output of the function is visualized on the middle frame of the series in Figure 5. It is important to note that the volume has been reduced to one channel at the fifth block, so it is not possible to generate a 3D Grad-CAM heatmap at this layer.

Across all images in Figure 5, it can be seen that the model focuses on the disc bulges as one of the main indicators of aging. The second image from the left shows a relatively healthy spine for a 77 year old, with a predicted spine age of 34.75 years. The third image is the series the model has attended to C6 as radiologists have suggested, but focused more on disc bulges. Based on radiology feedback, the spine age should be younger than the chronological age but around 45 to 50 years, not 39 years. The last picture depicts a participant whose spine has degenerated significantly compared to their chronological age.

### 4.5 Spine Age Gap as a Biomarker for Spine Health

In this section, we explore the relationship between spine conditions, lifestyle factors, and predicted spine age. Specifically, we aim to determine whether the gap between chronological age and predicted spine age is associated with clinically relevant spine conditions, and whether this spine age gap can be used as a biomarker for overall spine health.

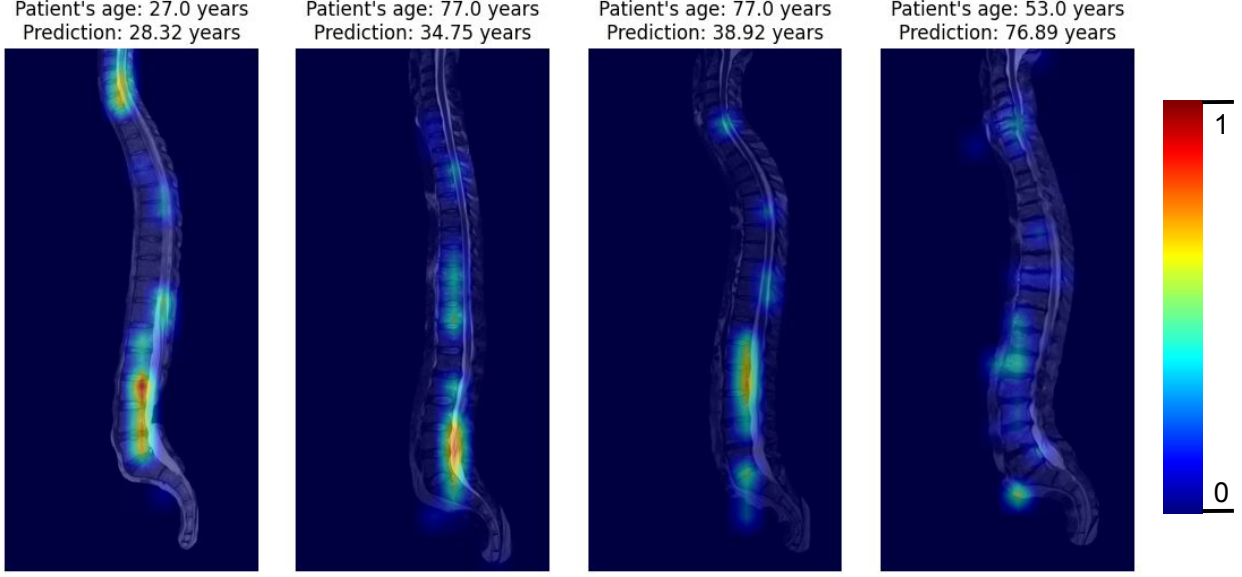


Figure 5: Grad-CAM heatmap on the middle frame of the MRI for four different patients. The values are adjusted based on  $f(x) = \max(\ln(288x), 1)$  to have better contrast.

We begin by quantifying the associations between SAG and spinal conditions in Table 5. We fit separate linear regression models using lumbar degenerative conditions, spinal structural and canal pathologies, or lifestyle factors as covariates, with SAG as the predicted outcome. Lumbar degenerative conditions were categorized and in conditions with multiple severities cases were only assigned to the most severe category present. We observe that lumbar disc bulges and osteophytes are significantly associated with SAG, with the occurrence of a severe disc bulge resulting in a predicted increased SAG of 2.96 years on average. Several spinal and canal pathologies are also associated with an increased SAG, including spinal stenosis, fractures, and spinal canal narrowing. We find similar associations between SAG and cervical degenerative conditions (see the Appendix for details).

Examining the association between SAG and lifestyle factors, we find higher levels of smoking and physically demanding work are associated with an increase in SAG, while vigorous exercise is associated with a decrease in SAG. This association with between spine health and smoking aligns with previous studies [44, 45] that identified an association between smoking and intervertebral disc degeneration, as well as changes in vertebral bone structure. Alcohol consumption has only a mild association with SAG, which again aligns with previous research indicating no substantial relationship between alcohol intake and spinal degeneration [46, 47, 48, 49].

Next, we explore whether large spine age gaps are associated with an increased likelihood of clinically relevant spinal conditions. We compare the odds of degenerative lumbar and spinal structural conditions in the test set of cases where individuals had a SAG of greater than 5, with cases where individuals had a SAG of less than -5 (Figure 6). We find the odds of individuals having lumbar disc bulges or osteophytes much higher in the group with  $SAG > 5$ . Individuals with a large positive SAG have odds of moderate disc bulges four-times higher than individuals with a large negative SAG, and this grows to eight-fold higher for severe disc bulges. Similarly, the odds of spinal fractures, stenosis, spondylolisthesis, and canal narrowing are two- to four-times larger in individuals with a large positive SAG.

The relationship between spine health and lifestyle factors may change over chronological age. For example, it may be that performing physically demanding work degrades spine health when individuals are young. However, for older individuals, the ability to perform physically demanding work at all may suggest better spine health relative to those not performing such work. This reverse in relationship over time is exactly what we see when comparing SAG and actual age in individuals who do and do not perform physically demanding work (Figure 7). Average SAG is higher for younger individuals performing physically demanding work, but lower for older individuals. This shows that SAG may be a relevant marker of overall spine health.

Table 5: Linear regression coefficients and 95% confidence intervals (CIs) quantifying the association between lumbar degenerative conditions, spine structural and canal pathologies, and lifestyle factors with spine-age gap. Separate regression models were fit for lumbar degenerate conditions, spinal structural pathologies, and lifestyle factors. In all cases models controlled for biological sex.

	Condition	n	Effect (95% CI)
Lumbar degenerative spinal conditions	Mild disc bulge (No.>2)	801	1.27 (0.92, 1.63)*
	Moderate disc bulge (No.>1)	85	1.58 (0.55, 2.61)*
	Severe disc bulge (No.>0)	12	2.96 (0.27, 5.65)*
	Mild disc dessication (No.>1)	315	-0.30 (-0.84, 0.25)
	Moderate disc dessication (No.>0)	295	0.02 (-0.55, 0.59)
	Severe disc dessication (No.>0)	212	0.20 (-0.48, 0.88)
	Near complete disc dessication (No.>0)	66	-0.59 (-1.76, 0.59)
	Disc annular fissure (No.>0)	781	0.24 (-0.12, 0.59)
	Vertebral endplate change (No.>0)	547	0.19 (-0.26, 0.64)
	Mild disc osteophyte (No.>1)	630	2.34 (1.95, 2.73)*
	Moderate disc osteophyte (No.>0)	36	2.46 (0.88, 4.03)*
	Mild disc protrusion (No.>1)	249	0.41 (-0.19, 1.02)
	Moderate disc protrusion (No.>0)	94	0.54 (-0.43, 1.51)
	Mild disc extrusion (No.>0)	71	-0.59 (-1.70, 0.52)
Spinal structural and canal pathologies	Spondylolisthesis	722	0.91 (0.56, 1.26)*
	Scoliosis	1477	0.64 (0.37, 0.91)*
	Kyphosis or lordosis	3061	0.45 (0.23, 0.66)*
	Fracture	229	1.45 (0.82, 2.07)*
	Spinal stenosis	125	1.87 (1.03, 2.72)*
	Congenital spinal canal narrowing	197	1.18 (0.50, 1.85)*
	Cord abnormalities	194	0.22 (-0.46, 0.90)
	Transitional vertebra	542	-0.07 (-0.49, 0.34)
	Tralov perineural cyst	601	-0.30 (-0.70, 0.10)
Lifestyle factors	Packs per day smoked ‡	2302	0.93 (0.64, 1.22)*
	Days per week consuming alcohol ‡	6136	0.08 (0.03, 0.13)*
	Time sedentary ‡	8332	-0.01 (-0.04, 0.03)
	Physically moderate work	1001	0.29 (-0.03, 0.62)
	Physically heavy work	488	0.67 (0.22, 1.12)*
	Moderate exercise	2146	-0.40 (-0.70, -0.11)*
	Vigorous exercise	4285	-0.79 (-1.05, -0.52)*

\* statistically significant effects. ‡ continuous variables, with counts representing the number greater than zero.

## 5 Conclusion

We believe this work is a significant step towards understanding spine health and its key factors. By analyzing radiology reports, we developed a data-driven framework for detecting typical age-related spine degradation in different age groups, assessed alternate approaches to achieve optimal model performance for predicting spine age, and demonstrated that SAG is an indicator of spine health based on its associations with spinal conditions and lifestyle factors.

This study is the first to utilize deep learning and convolutional neural networks to estimate spine age based on sagittal T2-weighted MRI series. We developed a network inspired by brain age estimation research to estimate spine age. Our experiments were conducted on a vast dataset of over 18,000 MRI series and 17,000 participants, with a detailed study on the clinical relevance of SAG and spine conditions in more than 8,500 MRI series.

One limitation of our work is the lack of data on rare or severe conditions as the scans were primarily for preventive health screening. Hospitals that deal with patients with these rare conditions may enhance model performance with additional data. For example, conditions like soft tissue edema and severe lumbar disc osteophyte were rare in our dataset, limiting our ability to analyze model performance on these conditions. More data may also help improve our robustness to spine fractures, which was missed in one of the failure cases.

Future work includes:

1. Investigating the relationship between performance improvement and clinical findings by using more data, especially data with rare or severe spine conditions.

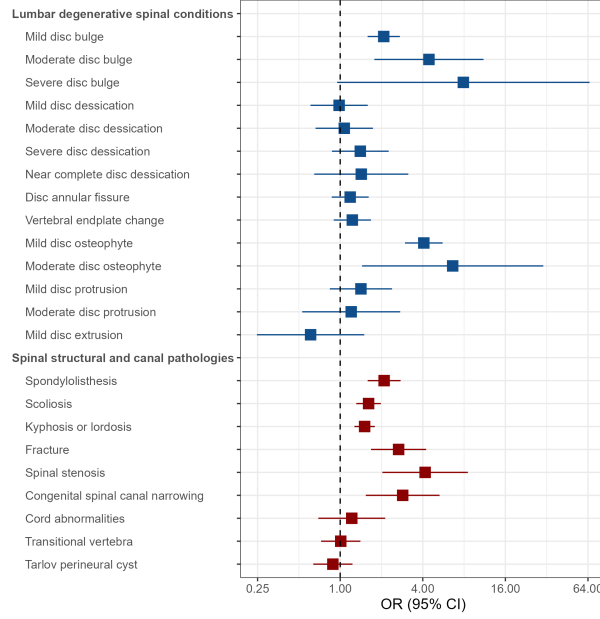


Figure 6: Odds ratios (OR) for lumbar degenerative and spinal structural conditions comparing cases with large positive spine age gaps (greater than 5 years) and large negative spine age gaps (less than -5 years). The x-axis depicts odds ratios between large and small spine age groups, with horizontal lines representing 95% confidence intervals for the ORs.

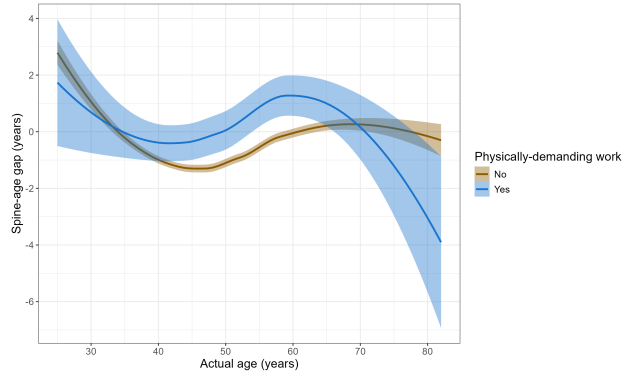


Figure 7: Relationship between spine-age gap and actual age in individuals who do and do not perform physically demanding work.

2. In our current model, we utilized a Deep Convolutional Neural Network. There is potential to investigate state-of-the-art deep learning models, such as vision transformers, which could enhance both the model performance and the clinical findings.
3. Additionally, we employed a combination of UMAP and HDBSCAN to identify normal and abnormal spine groups within age brackets. Future research could explore more effective methods to replace these techniques. For example, an encoder-decoder-based approach could be used for dimensionality reduction instead of UMAP, providing more meaningful features and potentially improving the identification of normal populations.
4. Furthermore, our technique of identifying normal participants, estimating age, and analyzing clinical relevance can be applied to other organs, such as prostate, kidney, and liver.

## References

- [1] Michael Papadakis, Georgios Sapkas, Elias C Papadopoulos, and Pavlos Katonis. Pathophysiology and biomechanics of the aging spine. *The open orthopaedics journal*, 5:335, 2011.
- [2] Key Yong Kim, Yung Tae Kim, Choon Sung Lee, and Myung Jin Shin. Mri classification of lumbar herniated intervertebral disc, 1992.
- [3] Christian WA Pfirrmann, Alexander Metzdorf, Marco Zanetti, Juerg Hodler, and Norbert Boos. Magnetic resonance classification of lumbar intervertebral disc degeneration. *spine*, 26(17):1873–1878, 2001.
- [4] Ron I Riesenburger, Mina G Safain, Richard Ogbuji, Jackson Hayes, and Steven W Hwang. A novel classification system of lumbar disc degeneration. *Journal of Clinical Neuroscience*, 22(2):346–351, 2015.
- [5] O Gille, V Challier, H Parent, R Cavagna, A Poignard, A Faline, S Fuentes, O Ricart, E Ferrero, M Ould Slimane, et al. Degenerative lumbar spondylolisthesis. cohort of 670 patients, and proposal of a new classification. *Orthopaedics & Traumatology: Surgery & Research*, 100(6):S311–S315, 2014.
- [6] Olivier Gille, Houssam Bouloussa, Simon Mazas, Claudio Vergari, Vincent Challier, Jean-Marc Vital, Pierre Coudert, and Soufiane Ghailane. A new classification system for degenerative spondylolisthesis of the lumbar spine. *European Spine Journal*, 26:3096–3105, 2017.
- [7] Jen-Tang Lu, Stefano Pedemonte, Bernardo Bizzo, Sean Doyle, Katherine P Andriole, Mark H Michalski, R Gilberto Gonzalez, and Stuart R Pomerantz. Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. In *Machine Learning for Healthcare Conference*, pages 403–419. PMLR, 2018.
- [8] Jiabo He, Wei Liu, Yu Wang, Xingjun Ma, and Xian-Sheng Hua. Spineone: A one-stage detection framework for degenerative discs and vertebrae. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1331–1334. IEEE, 2021.
- [9] James Thomas Patrick Decourcy Hallinan, Lei Zhu, Kaiyuan Yang, Andrew Makmur, Diyaa Abdul Rauf Algazwi, Yee Liang Thian, Samuel Lau, Yun Song Choo, Sterling Ellis Eide, Qai Ven Yap, et al. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine mri. *Radiology*, 300(1):130–138, 2021.
- [10] Hua-Dong Zheng, Yue-Li Sun, De-Wei Kong, Meng-Chen Yin, Jiang Chen, Yong-Peng Lin, Xue-Feng Ma, Hong-Shen Wang, Guang-Jie Yuan, Min Yao, et al. Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from mri. *Nature communications*, 13(1):841, 2022.
- [11] Wei Yi, Jingwei Zhao, Wen Tang, Hongkun Yin, Lifeng Yu, Yaohui Wang, and Wei Tian. Deep learning-based high-accuracy detection for lumbar and cervical degenerative disease on t2-weighted mr images. *European Spine Journal*, 32(11):3807–3814, 2023.
- [12] Kaisi Chen, Lei Zheng, Honghao Zhao, and Zihang Wang. Deep learning-based intelligent diagnosis of lumbar diseases with multi-angle view of intervertebral disc. *Mathematics*, 12(13):2062, 2024.
- [13] Wei Shen, Yilu Guo, Yan Wang, Kai Zhao, Bo Wang, and Alan L Yuille. Deep regression forests for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2304–2313, 2018.
- [14] Zhifei Zhang, Yang Song, and Hairong Qi. Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818, 2017.
- [15] Xiangbo Shu, Jinhui Tang, Hanjiang Lai, Luoqi Liu, and Shuicheng Yan. Personalized age progression with aging dictionary. In *Proceedings of the IEEE international conference on computer vision*, pages 3970–3978, 2015.
- [16] Ira Kemelmacher-Shlizerman, Supasorn Suwajanakorn, and Steven M Seitz. Illumination-aware age progression. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3334–3341, 2014.
- [17] Biao Zhang, Shuqin Zhang, Jianfeng Feng, and Shihua Zhang. Age-level bias correction in brain age prediction. *NeuroImage: Clinical*, 37:103319, 2023.
- [18] James H Cole, Stuart J Ritchie, Mark E Bastin, Valdés Hernández, S Muñoz Maniega, Natalie Royle, Janie Corley, Alison Pattie, Sarah E Harris, Qian Zhang, et al. Brain age predicts mortality. *Molecular psychiatry*, 23(5):1385–1392, 2018.
- [19] Jay Shah, Md Mahfuzur Rahman Siddiquee, Yi Su, Teresa Wu, and Baoxin Li. Ordinal classification with distance regularization for robust brain age prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7882–7891, 2024.

- [20] Liyue Shen, Jimmy Zheng, Edward H Lee, Katie Shpanskaya, Emily S McKenna, Mahesh G Atluri, Dinko Plasto, Courtney Mitchell, Lillian M Lai, Carolina V Guimaraes, et al. Attention-guided deep learning for gestational age prediction using fetal brain MRI. *Scientific reports*, 12(1):1408, 2022.
- [21] Benedikt Atli Jónsson, Gyda Bjornsdottir, TE Thorgeirsson, Lotta María Ellingsen, G Bragi Walters, DF Gudbjartsson, Hreinn Stefansson, Kari Stefansson, and MO Ulfarsson. Brain age prediction using deep learning uncovers associated sequence variants. *Nature communications*, 10(1):5409, 2019.
- [22] Jeyeon Lee, Brian J Burkett, Hoon-Ki Min, Matthew L Senjem, Emily S Lundt, Hugo Botha, Jonathan Graff-Radford, Leland R Barnard, Jeffrey L Gunter, Christopher G Schwarz, et al. Deep learning-based brain age prediction in normal aging and dementia. *Nature Aging*, 2(5):412–424, 2022.
- [23] Ann-Marie G de Lange, Melis Anatürk, Jaroslav Rokicki, Laura KM Han, Katja Franke, Dag Alnæs, Klaus P Ebmeier, Bogdan Draganski, Tobias Kaufmann, Lars T Westlye, et al. Mind the gap: Performance metric evaluation in brain-age prediction. *Human Brain Mapping*, 43(10):3113–3129, 2022.
- [24] Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical image analysis*, 68:101871, 2021.
- [25] Karim Armanious, Sherif Abdulatif, Wenbin Shi, Shashank Salian, Thomas Küstner, Daniel Weiskopf, Tobias Hepp, Sergios Gatidis, and Bin Yang. Age-net: An MRI-based iterative framework for brain biological age estimation. *IEEE Transactions on Medical Imaging*, 40(7):1778–1791, 2021.
- [26] Jian Cheng, Ziyang Liu, Hao Guan, Zhenzhou Wu, Haogang Zhu, Jiyang Jiang, Wei Wen, Dacheng Tao, and Tao Liu. Brain age estimation from MRI using cascade networks with ranking loss. *IEEE Transactions on Medical Imaging*, 40(12):3400–3412, 2021.
- [27] Zhaonian Zhang, Richard Jiang, Ce Zhang, Bryan Williams, Ziping Jiang, Chang-Tsun Li, Paul Chazot, Nicola Pavese, Ahmed Bouridane, and Azeddine Beghdadi. Robust brain age estimation based on sMRI via nonlinear age-adaptive ensemble learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 30:2146–2156, 2022.
- [28] James H Cole, Rudra PK Poudel, Dimosthenis Tsagkrasoulis, Matthan WA Caan, Claire Steves, Tim D Spector, and Giovanni Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163:115–124, 2017.
- [29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [30] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [31] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [32] Satoshi Watanabe and Koichi Terazawa. Age estimation from the degree of osteophyte formation of vertebral columns in japanese. *Legal medicine*, 8(3):156–160, 2006.
- [33] Frank J Rühli, M Müntener, and M Henneberg. Age-dependent changes of the normal human spine during adulthood. *American Journal of Human Biology: The Official Journal of the Human Biology Association*, 17(4):460–469, 2005.
- [34] RA Savage, GH Whitehouse, and N Roberts. The relationship between the magnetic resonance imaging appearance of the lumbar spine and low back pain, age and occupation in males. *European Spine Journal*, 6:106–114, 1997.
- [35] Atif Khan, Daciana Iliescu, Evor Hines, Charles Hutchinson, and Robert Sneath. Neural network based spinal age estimation using lumbar spine magnetic resonance images (MRI). In *2013 4th International Conference on Intelligent Systems, Modelling and Simulation*, pages 88–93. IEEE, 2013.
- [36] Robert JS Sneath, Atif Khan, and Charles Hutchinson. An objective assessment of lumbar spine degeneration/ageing seen on MRI using an ensemble method—a novel approach to lumbar MRI reporting. *Spine*, 47(5):E187–E195, 2022.
- [37] Siavash Khallaghi, Lucas Porto, Sean London, Yosef Chodakiewitz, Rajpaul Attariwal, and Sam Hashemi. Quantitative assessment of the whole spine in T2 MRI using deep learning. *International Society for Magnetic Resonance in Medicine (ISMRM)*, 2023.
- [38] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

- [39] Leland McInnes, John Healy, Steve Astels, et al. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205, 2017.
- [40] EH Leonardsen, H Peng, T Kaufmann, I Agartz, OA Andreassen, EG Celius, T Espeseth, HF Harbo, EA Høgestøl, AM de Lange, et al. Deep neural networks learn general and clinically relevant representations of the ageing brain. *neuroimage*, 256, article 119210, 2022.
- [41] Soojin Lee, Saurabh Garg, Madhurima Datta, Duc Nguyen, Nasrin Akbari, Arun Rajendran, Sam Hashemi, Yosef Chodakiewitz, and Rajpaul Attariwala. Smoking is associated with accelerated brain aging: A large, diverse population-based analysis (s43. 003). In *Neurology*, volume 102, page 2376. AAN Enterprises, 2024.
- [42] Charlotte Leboeuf-Yde, Jan Nielsen, Kirsten O Kyvik, René Fejer, and Jan Hartvigsen. Pain in the lumbar, thoracic or cervical regions: do age and gender matter? a population-based study of 34,902 Danish twins 20–71 years of age. *BMC musculoskeletal disorders*, 10:1–12, 2009.
- [43] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [44] Mohammed Akmal, Anil Kesani, Bobby Anand, Abhinav Singh, Mike Wiseman, and Allen Goodship. Effect of nicotine on spinal disc cells: a cellular mechanism for disc degeneration. *Spine*, 29(5):568–575, 2004.
- [45] Daniel Berman, Jonathan H Oren, John Bendo, and Jeffrey Spivak. The effect of smoking on spinal fusion. *International journal of spine surgery*, 11(4), 2017.
- [46] DM Gorman, G Potamianos, KA Williams, AO Frank, SW Duffy, and TJ Peters. Relationship between alcohol abuse and low back pain. *Alcohol and Alcoholism*, 22(1):61–63, 1987.
- [47] Charlotte Leboeuf-Yde. Alcohol and low-back pain: a systematic literature review. *Journal of Manipulative and Physiological Therapeutics*, 23(5):343–346, 2000.
- [48] Sara AC Holmberg and Anders G Thelin. Primary care consultation, hospital admission, sick leave and disability pension owing to neck and low back pain: a 12-year prospective cohort study in a rural population. *BMC Musculoskeletal Disorders*, 7:1–8, 2006.
- [49] Yanwei Lv, Wei Tian, Dafang Chen, Yajun Liu, Lifang Wang, and Fangfang Duan. The prevalence and associated factors of symptomatic cervical spondylosis in chinese adults: a community-based cross-sectional study. *BMC musculoskeletal disorders*, 19:1–12, 2018.



## Appendix

Table 6: Linear regression coefficients and 95% confidence intervals (CIs) quantifying the association between cervical and thoracic degenerative conditions with spine-age gap. Separate regression models were fit for cervical and thoracic degenerate conditions. In all cases models controlled for biological sex.

	Condition	n	Effect (95% CI)
Cervical degenerative spinal conditions	Mild disc bulge (No.>2)	459	0.64 (0.19, 1.10)*
	Moderate disc bulge (No.>1)	32	2.26 (0.60, 3.92)*
	Severe disc bulge (No.>0)	9	1.09 (-2.03, 4.21)
	Mild disc dessication (No.>1)	243	-0.23 (-0.85, 0.38)
	Moderate disc dessication (No.>0)	210	0.44 (-0.23, 1.10)
	Severe disc dessication (No.>0)	124	-0.03 (-0.89, 0.84)
	Near complete disc dessication (No.>0)	27	-1.14 (-2.94, 0.67)
	Disc annular fissure (No.>0)	79	0.18 (-0.88, 1.25)
	Vertebral endplate change (No.>0)	208	0.06 (-0.61, 0.74)
	Mild disc osteophyte (No.>1)	1805	1.12 (0.86, 1.38)*
	Moderate disc osteophyte (No.>0)	276	1.09 (0.51, 1.67)*
	Mild disc protrusion (No.>1)	271	0.43 (-0.15, 1.01)
	Moderate disc protrusion (No.>0)	64	1.00 (-0.17, 2.18)
	Mild disc extrusion (No.>0)	12	-1.44 (-4.15, 1.26)
Thoracic degenerative spinal conditions	Mild disc bulge (No.>2)	136	0.75 (-0.07, 1.58)
	Mild disc dessication (No.>1)	96	-0.04 (-1.03, 0.94)
	Moderate disc dessication (No.>0)	43	-1.01 (-2.44, 0.43)
	Severe disc dessication (No.>0)	15	-3.42 (-5.87, -0.96)*
	Disc annular fissure (No.>0)	24	0.63 (-1.30, 2.56)
	Vertebral endplate change (No.>0)	124	0.78 (-0.09, 1.65)
	Mild disc osteophyte (No.>1)	30	1.13 (-0.59, 2.85)
	Mild disc protrusion (No.>1)	177	0.67 (-0.04, 1.39)
	Moderate disc protrusion (No.>0)	26	1.82 (-0.03, 3.66)
	Mild disc extrusion (No.>0)	24	-1.72 (-3.64, 0.20)

\* statistically significant effects.