ProgreSpine: Inherently Explainable Prototypical Regression for Spine Age Estimation

Roozbeh Bazargani[1[0009−0001−7197−2953]], Saqib Basar[1[0000−0002−8596−1040]], Sam Hashemi[1[0009−0003−0640−6466]], and Siavash Khallaghi[1[0000−0003−2543−1934]]

Prenuvo
roozbeh.bazargani@prenuvo.com

Abstract. Spine aging is a complicated process shaped by pathologies, genetic factors, and lifestyle influences. Radiologists routinely use MR images to assess the spinal health of patients in different age brackets. Quantifying spinal health as an organ age would allow ranking and monitoring of patients within the same and across different demographics. However, spine age estimation has been limited to classical machine learning methods which suffer from high error rates and a lack of interpretability. Moreover, inherently explainable state-of-the-art models in organ age estimation, such as prototypical networks, are limited to 2D and are not extendable to repeated prototype labels. This is important as organs typically degenerate in different ways as a result of aging. We propose ProgreSpine, the first deep-learning-based 3D spine age estimation model based on prototypical regression with a loss specifically tailored to repeated prototype labels. We trained and tuned our proposed model on a large dataset of 9542 samples and performed a thorough evaluation on 1069 samples to demonstrate improved performance against the state-of-the-art with a mean absolute error of 3.61 years. Furthermore, the results suggest that the model learns the prototypes based on clinical conditions that will facilitate monitoring disease progression with a transparent model. The source code is available at https://github.com/prenuvo/progrespine.

Keywords: Prototypical Networks · Age Estimation · Spine Degeneration · Explainable AI · MRI

1 Introduction

Spine aging is a multifaceted process influenced by pathologies, genetics, and lifestyle factors [19]. Magnetic Resonance Imaging (MRI) offers excellent visualization of the musculoskeletal system, allowing the assessment of age-related morphological changes. Historically, radiologists have used this modality to develop protocols for patient stratification and holistic assessment of spine degeneration [8,21]. The assignment of a biological age to the spine that correlates with age-related degradation provides patients and radiologists with a reliable indicator of spinal health. It enables more accurate risk stratification by identifying patients with advanced degeneration relative to their chronological age.

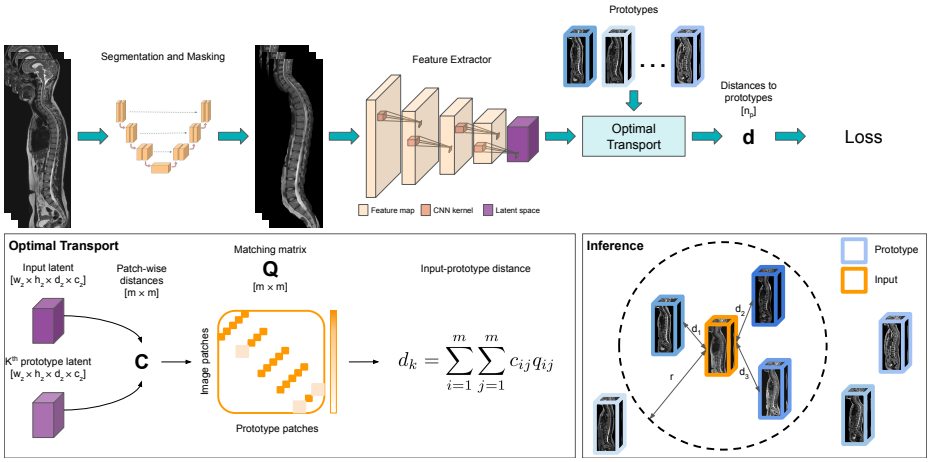

Fig. 1. Our model predicts spine age based on the distance of the input 3D MRI to the prototypes from the training set. In the beginning, the background organs are masked using a segmentation model. Next, Prototypes are learned by training a 3D CNN feature extractor. The prototype set is the subset of the training set and each prototype represents a whole spine image. The predicted age is computed as the weighted average of the prototype labels based on their distance.

It can also inform clinical decision-making by assessing disease progression and provide patients with an individualized measure to guide long-term care.

To the best of our knowledge, deep learning approaches have not been explored for spine age estimation and most recent studies rely on classical machine learning such as random forests, extreme gradient boosting trees, and support vector machines [14,25]. However, these methods were limited by small datasets ($<$100), relied on manually extracted features and exhibited poor performance with a Mean Absolute Error (MAE) of 10.28 years [25]. Given the feasibility of detecting spine degeneration using deep learning [5,9,10,16,29,28], it might be possible to address the aforementioned shortcomings in classical spine age estimation through end-to-end deep learning.

Deep-learning-based age estimation is typically achieved through either regression or binning. In the regression approach, the model directly predicts a continuous value representing the chronological age [3,6,24]. In the binning approach, the chronological age is rounded to the nearest integer and treated as a categorical variable. Subsequently, the model predicts the probability of organ age belonging to each bin [20,23]. However, despite their effectiveness, these studies rely on black-box architectures that offer little insight into the decision-making process which impedes clinical trust and regulatory approval. By incorporating Explainable Artificial Intelligence (XAI) techniques, we can illuminate which features and regions contribute most significantly to age estimation.

XAI comprises post-hoc and inherently explainable (ante-hoc) methods. Post-hoc methods attempt to explain black box model decisions after training, often

using saliency techniques to highlight the influence of individual input voxels on predictions. However, these explanations do not reliably reflect the model's decision-making process. The resulting maps are often noisy, non-robust, and misleading, making it challenging to explain why the model favors one class over another [1,2,11,12,22,27].

Inherently explainable techniques include prototypical networks. The architecture of these models shows how much each prototype from the training set has contributed to the final prediction. Several studies have used prototypical networks for classification in healthcare applications [11,12,15,17,27,30]. The only regression studies in medical imaging are INSightR-Net [12] and ExPeRT [11]. INSightR-Net is an ordinal regression task on 2D images of eye without a truly continuous latent space [11]. Hesse et al. [11] proposed ExPeRt which estimates brain age using a single 2D MRI slice. Their proposed model outperforms INSightR-Net in terms of MAE.

We have introduced ProgreSpine a prototypical network to regress spine age based on T2-weighted whole-spine MR images. This model takes 3D images as input, segments the spine as the region of interest, and estimates spine age based on the distance between the input and prototype embeddings. As spine degeneration includes curvature disorders and might include multiple regions, unlike patch-based prototypical networks [12,27] and in alignment with ExPeRT [11], we decided to use the entire image as the prototype. This work improves upon ExPeRT for 2D age estimation [11] by extending the model from 2D to 3D MRI. We introduced a new loss to handle several prototypes with the same age label as spine degeneration might have different spondylosis appearances in T2 MRI [21] and one sample might not necessarily represent all degeneration types.

Our contributions, to the best of our knowledge, are the following. This manuscript is the first 3D extension of prototypical regression. We also present the first deep-learning approach to spine age estimation. Furthermore, we tailor the prototypical loss to handle different types of degeneration. Finally, we perform an extensive analysis of age estimation methods on a large dataset of 10,611 whole-spine MRI with 1,069 samples in the testing set.

2 Method

Figure 1 shows an overview of our proposed approach. The field of view in a whole spine MRI spans multiple organs that age at different rates. To disentangle the spine from the rest of the organs, we use a U-Net semantic segmentation model similar to Khallaghi et al. [13]. 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. To decrease the spatial variability of samples in our dataset, we resample all series to a common spacing of $0.9 \times 0.9 \times 3\ mm^3$. Subsequently, we center-cropped or padded all images to a fixed size of $384 \times 793 \times 14$.

2.1 Prototypical Network Architecture

Our goal is to train a fully convolutional network based on the set of 3D MR images $X \subset \mathbb{R}^{w \times h \times d}$ to extract the latent space set $Z \subset \mathbb{R}^{w_z \times h_z \times d_z \times c_z}$ where w_z, h_z, d_z, and c_z represent width, height, depth, and channel dimension. The backbone of the feature extractor consists of five 3D convolution blocks and a top block with a sigmoid activation function inspired by Peng et al. [20]. Concurrently, we learn a set of prototypes $P = \cup_{y \in Y} P_y$ where $Y \subset \mathbb{R}$ is the set of prototype labels and $P_y \subset \mathbb{R}^{w_z \times h_z \times d_z \times c_z}$ is the set of prototypes belonging to label y. $|P| = n_p$ and $|Y| = n_y$ where n_p and n_y are the total number of prototypes and unique labels of prototypes, respectively. The vector $y^p \in \mathbb{R}^{n_p}$ denotes prototype labels. Labels are set at the start of training and remain fixed throughout.

The prototypes are part of the model parameters and are simultaneously learned with the feature extractor. Therefore, after each iteration, the prototypes are updated. However, the prototypes do not necessarily indicate an image that can be visualized at each step. Therefore, we have to project the prototypes to the closest image representation in the latent space. Projection is done every N epochs similar to ProtoPNet [4].

2.2 Training Objective

As the first step, we need to define the distance between the image representations in the latent space. Since latent space forms a high-dimensional manifold, measuring feature distances is challenging [11]. In the local neighborhood, this distance is approximated by Euclidean distance, as demonstrated by Teenbaum et al. [26], who showed that Euclidean distance effectively approximates small-scale distances on manifolds.

To compute the distance between the two image representations, one needs to recall that the latent space $Z \in \mathbb{R}^{w_z \times h_z \times d_z \times c_z}$ consists of $m = w_z h_z d_z$ patches. We define the pair-wise patch distance matrix $C \in \mathbb{R}^{m \times m}, c_{ij} = |z_i - p_j|^2$ where z_i and p_j are the embedding vectors of size c_z of i-th patch of an image representation and j-th patch of one of the prototypes. We opt for optimal transport (OT) [11] that would define d_k, the distance of the input image representation to the k-th prototype, based on C using a matching matrix $Q \in \mathbb{R}^{m \times m}$:

$$d_k = \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} q_{ij} \tag{1}$$

where c_{ij} and q_{ij} are elements of C and Q. OT is especially important in this context, given that the spine might not be centered in the volume and its curvature might exhibit abnormalities such as lordosis, kyphosis or scoliosis. As a result, a specific part of the spine in one image might correspond to the i-th patch in one image and j-th patch in another. OT ensures that these patches are compared against each other.

In order to employ OT, we define the following optimization problem to minimize the distance Q with constraints:

$$\min_{Q} \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} q_{ij} \text{ s.t. } \sum_{i=1}^{m} q_{ij} = w_1, \sum_{j=1}^{m} q_{ij} = w_2 \tag{2}$$

where w_1 and w_2 are the initial marginal distributions. However, this is a computationally expensive problem. Cuturi [7] introduced the entropic regularization to smooth the optimization problem and transformed Eq. 2 into:

$$\min_{Q} \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} q_{ij} + \epsilon H(Q), H(Q) = -\sum_{i=1}^{m} \sum_{j=1}^{m} q_{ij} log(q_{ij}) \tag{3}$$

where ϵ is the regularization weight. This objective is solvable using the classical Sinkhorn divergence algorithm and can be used to train the network end-to-end as it is fully differentiable [11].

d'_k is defined as the softmin weighted average of distances to P_{y_k}:

$$d'_k = \sum_{d_i \in D_{y_k}} \frac{d_i . e^{-d_i}}{\sum_{d_j \in D_{y_k}} e^{-d_j}} \tag{4}$$

where D_{y_k} is the list of distances to prototypes with label y_k. We aim to regularize the distances between samples and prototypes according to their label (age) differences: $d'_k \propto |y_k - y|$ where y is the sample label and y_k is the k-th prototype group label. To this end, we define the loss for a sample as:

$$L(d, Y, y) = \sum_{y_k \in Y} (|s.d'_k - (|y - y_k|)|) w_k^{train} \tag{5}$$

where s is a learnable scaling parameter and w_k^{train} is the weight associated with the k-th prototype group sample defined as:

$$w_k^{train} = e^{-\frac{(y-y_k)^2}{2\sigma^2}} + \alpha \tag{6}$$

where σ is the standard deviation of the Gaussian function that controls the local neighborhood size and α is a small number that ensures samples and prototypes with a large label difference have a contribution to the loss and stay far from each other, preventing their collapse. The idea behind the softmin is to only force the closest prototype in each age to be regularized by the distance, and the rest of the prototypes can learn other types of degeneration. Moreover, softmin is preferred over the minimum as it allows gradients to flow for all prototypes. Otherwise, some prototypes may be ignored and never updated during training.

2.3 Inference

To predict the age of a new spine MRI, we perform a weighted average on the prototype labels in distance r from the inference image representation in the

latent space. The weights are defined using a Gaussian kernel:

$$\hat{y} = \frac{\sum_{k=1}^{n_p} w_k^{test} y_k^p}{\sum_{k=1}^{n_p} w_k^{test}}, w_k^{test} = \begin{cases} e^{-\frac{(s.d_k)^2}{2(r/3)^2}}, & \text{if } s.d_k \leq r \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where y_k^p is the k-th prototype label. After training, we used the validation data to correct the bias towards the mean using Peng et al. method [20].

3 Experiments and Results

3.1 Data

We utilized a comprehensive dataset comprising of 10,611 3D T2-weighted whole spine MRI series with the sagittal view as the imaging plane. These images are reported normal in appearance in the radiology report. The scans were collected from 2011 to 2024, using 19 scanner machines consisting of Siemens Magnetom Aera, Siemens Espree, and Philips Ingenia Ambition X across 10 clinics in North America. The dataset included individuals aged 25 to 84 years, including those with variational anatomy. The data was divided based on age and gender into training, validation, and testing sets of 8491, 1051, and 1069 samples.

3.2 Implementation

For the experiments, we set the prototype labels to have a gap of 2 years $n_y = 31$ and be repeated 3 times for each age, summing up to $n_p = 93$ prototypes. The dimensions of the latent space Z were $w_z = 12$, $h_z = 24$, $d_z = 1$, and $c_z = 64$. The network parameters were selected or initialized as follows: $r = 5$, $\lambda = 0.1$, $s = 10$, $\sigma = 1$, $\alpha = 0.2$. The learning rates were set to 0.0005 (divided by half every 10 epochs) and 0.01 for network parameters and the scaling parameter, respectively. We trained the model for 50 epochs and every $N = 5$ epochs projected the prototypes. Batch size of 2 was used and we accumulated the gradients for 3 iterations before backpropagation. We used an instance with an Nvidia A10G GPU to train our model.

3.3 Quantitative Analysis

A comparison of performance with previous work and ablation study based on MAE and R^2 is shown in Table 1. We compared our model against the Simple Fully Convolutional Neural Network (SFCN) [20] feature extractor (similar to our feature extractor) with Mean Squared Error (MSE) and ordinal [18] losses. We also compared against the prototypical-based method for age estimation, ExPeRT [11], but extended the feature extractor to 3D (3D-ExPeRT) with the same architecture as our proposed model. We tested the 3D-ExPeRT with and without Repeated Prototype (RP) labels. For 3D-ExPeRT without RP, we used 60 prototypes, one sample per age from 25 to 84.

Table 1. Comparison with previous work and an ablation study after bias correction. The best performance is shown in **bold**, while the second-best is underlined.

	OT	RP	MAE [yr]	R^2
SFCN [20] + MSE	-	-	3.90	0.833
SFCN [20] + Ordinal [18]	-	-	<u>3.83</u>	0.839
3D-ExPeRT [11]	✓	×	4.44	0.787
3D-ExPeRT [11]	✓	✓	4.10	0.822
ProgreSpine	×	✓	3.85	<u>0.841</u>
ProgreSpine	✓	×	4.61	0.763
ProgreSpine	✓	✓	**3.61**	**0.857**

The best state-of-the-art (SOTA) model was SFCN + Ordinal loss. The proposed model improved the performance by 0.22 years in MAE and 0.018 in R^2. Compared to the SOTA explainable model, i.e. 3D-ExPeRT without RP, our model still improved the performance by 0.83 years and 0.07 in R^2, highlighting the importance of repeated prototypes and the proposed loss. We also tried the repeated prototype labels for 3D-ExPeRT. It can be seen that ExPeRT loss is not generalizable to repeated prototypes as it tries to regularize the distances to all of the prototypes of the same label simultaneously. This limits model flexibility in learning different types of prototypes (degeneration) for the same age.

We also performed an ablation study on OT and RP. We replaced OT with average pooling, where we took the average patch embeddings and computed the Euclidean distance between the input and the prototypes. We observed that OT outperformed average pooling, indicating the importance of patch matching in computing the distance to prototypes. Finally, adding RP to ProgreSpine improved performance in terms of MAE and R^2. This suggests including more prototypes in each age group allows for capturing different types of spine degeneration which in turn leads to better performance and model explainability.

3.4 Qualitative Analysis

Figure 2 depicts the prototypes diversity across different clusters based on radiology reports. After embedding corresponding radiology report conditions into a vector and UMAP visualization, we found that prototypes cover all clusters, representing different degenerative patterns. Figure 3 demonstrates our inference results. A 65 years-old patient as input has a distance of 1.78 to prototype 58 (label: 63), 3.51 to prototype 59 (label: 63), and 38.24 to prototype 3 (label: 25). Prototype 58 is the closest prototype to the input. Both have lordosis, a straightened degenerated cervical section, and endplate change and multiple disc bulges in the lumbar region. However, the input has scoliosis and the prototype has a moderate change in a lumbar vertebra that causes the distance. Prototype 59 has the same label as 58, however, it has a healthier cervical without lordosis but with a bone lesion in the lumbar area. This suggests that the model does not suffer from prototype collapse for prototypes with the same label. The rest of the lumbar conditions of prototype 59 are similar to the input image and hence

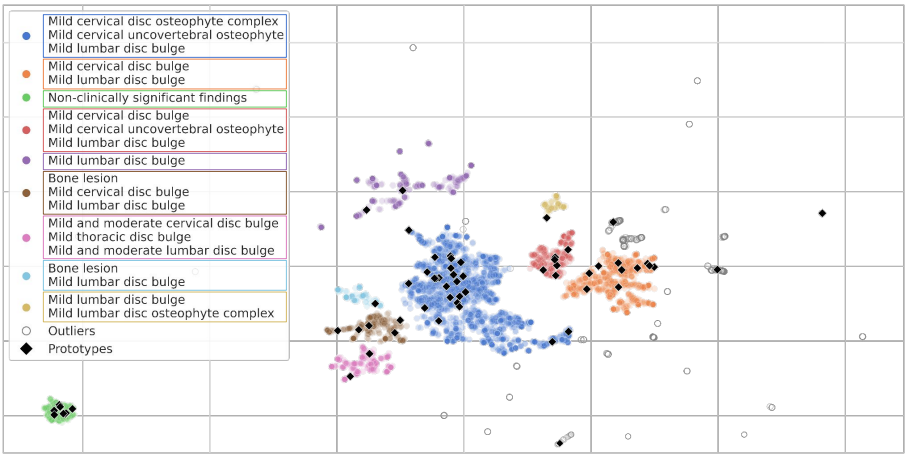

Fig. 2. Prototype diversity across clusters of the population based on radiology report.

they are not very far from each other. The farthest prototype is prototype 3 which represents a relatively healthy spine with a few minor lumbar disc bulges.

The weights are computed based on Eq. 7 and since the distance to prototype 80 is more than $r = 5$, the weight is 0. The regression plot of estimated spine age vs. chronological age is shown in Fig. 3 (top right). It can be seen that for individuals older than 80 years, ProgreSpine under-predicts. Notably, all baseline methods also under-predict in this range. This could be due to sparsity in the senior age bracket in our dataset, only 25 subjects aged over 80 out of 10611 (0.24%). The Bland-Altman plot demonstrates that the model has minimal bias (–0.05 years), indicating no systematic over- or underestimation. The limits of agreement (± 1.96 standard deviation) span approximately ± 9 years, which reflects not only model uncertainty but also potential biological variation where chronological age may not align with spine-specific aging. The bottom right figure shows prototypical distances, weights, and their labels for the input inference image. We observed that considering the minimum distance from the same label, the distances are regularized in alignment with differences in age.

4 Conclusion

We proposed ProgreSpine, the first deep-learning and prototypical regression method for spine age estimation. We extended the prototypical regression to 3D. Our results suggest the importance of having repeated prototypes since the spine degenerates in different ways. We generalized the prototypical loss [11] to repeated prototype labels and demonstrated improved performance. Finally, we used a large dataset of T2-weighted whole-spine to extensively explore SOTA models in spine age estimation. We observed that prototypes with similar conditions are closer in distance to the input image. Future work includes improving

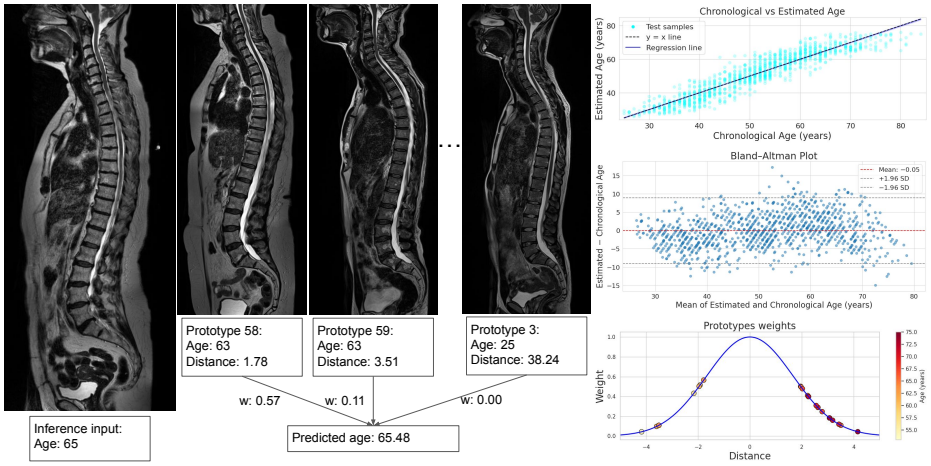

Fig. 3. Overview of the inference process. Left: The input image and its closest prototypes, showing their distances, ages, and contribution weights. Top right: Predicted vs. chronological age, and Bland–Altman plot illustrating agreement after bias correction. Bottom right: Weighting function applied to prototypes based on distance.

the model performance in the senior group (80+) with targeted data collection, extending this model to different organs, and exploring the potential of predicted spine age as a biomarker and its relation to spine conditions.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., Kim, B.: Sanity checks for saliency maps. Advances in neural information processing systems **31** (2018)
2. Alvarez-Melis, D., Jaakkola, T.S.: On the robustness of interpretability methods. arXiv preprint arXiv:1806.08049 (2018)
3. Armanious, K., Abdulatif, S., Shi, W., Salian, S., Küstner, T., Weiskopf, D., Hepp, T., Gatidis, S., Yang, B.: Age-net: An MRI-based iterative framework for brain biological age estimation. IEEE Transactions on Medical Imaging **40**(7), 1778–1791 (2021)
4. Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., Su, J.K.: This looks like that: deep learning for interpretable image recognition. Advances in neural information processing systems **32** (2019)
5. Chen, K., Zheng, L., Zhao, H., Wang, Z.: Deep learning-based intelligent diagnosis of lumbar diseases with multi-angle view of intervertebral disc. Mathematics **12**(13), 2062 (2024)
6. Cheng, J., Liu, Z., Guan, H., Wu, Z., Zhu, H., Jiang, J., Wen, W., Tao, D., Liu, T.: Brain age estimation from MRI using cascade networks with ranking loss. IEEE Transactions on Medical Imaging **40**(12), 3400–3412 (2021)

7. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems **26** (2013)
8. Gille, O., Bouloussa, H., Mazas, S., Vergari, C., Challier, V., Vital, J.M., Coudert, P., Ghailane, S.: A new classification system for degenerative spondylolisthesis of the lumbar spine. European Spine Journal **26**, 3096–3105 (2017)
9. Hallinan, J.T.P.D., Zhu, L., Yang, K., Makmur, A., Algazwi, D.A.R., Thian, Y.L., Lau, S., Choo, Y.S., Eide, S.E., Yap, Q.V., 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. He, J., Liu, W., Wang, Y., Ma, X., Hua, X.S.: Spineone: A one-stage detection framework for degenerative discs and vertebrae. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 1331–1334. IEEE (2021)
11. Hesse, L.S., Dinsdale, N.K., Namburete, A.I.: Prototype learning for explainable brain age prediction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7903–7913 (2024)
12. Hesse, L.S., Namburete, A.I.: Insightr-net: interpretable neural network for regression using similarity-based comparisons to prototypical examples. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 502–511. Springer (2022)
13. Khallaghi, S., Porto, L., London, S., Chodakiewitz, Y., Attariwal, R., Hashemi, S.: Quantitative assessment of the whole spine in T2 MRI using deep learning. International Society for Magnetic Resonance in Medicine (ISMRM) (2023)
14. Khan, A., Iliescu, D., Hines, E., Hutchinson, C., Sneath, R.: Neural network based spinal age estimation using lumbar spine magnetic resonance images (MRI). In: 2013 4th International Conference on Intelligent Systems, Modelling and Simulation. pp. 88–93. IEEE (2013)
15. Liang, Y., Liu, B., Zhang, H.: A convolutional neural network combined with prototype learning framework for brain functional network classification of autism spectrum disorder. IEEE Transactions on Neural Systems and Rehabilitation Engineering **29**, 2193–2202 (2021)
16. Lu, J.T., Pedemonte, S., Bizzo, B., Doyle, S., Andriole, K.P., Michalski, M.H., Gonzalez, R.G., Pomerantz, S.R.: Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. In: Machine Learning for Healthcare Conference. pp. 403–419. PMLR (2018)
17. Mulyadi, A.W., Jung, W., Oh, K., Yoon, J.S., Lee, K.H., Suk, H.I.: Estimating explainable Alzheimer's disease likelihood map via clinically-guided prototype learning. NeuroImage **273**, 120073 (2023)
18. Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G.: Ordinal regression with multiple output CNN for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4920–4928 (2016)
19. Papadakis, M., Sapkas, G., Papadopoulos, E.C., Katonis, P.: Pathophysiology and biomechanics of the aging spine. The open orthopaedics journal **5**, 335 (2011)
20. Peng, H., Gong, W., Beckmann, C.F., Vedaldi, A., Smith, S.M.: Accurate brain age prediction with lightweight deep neural networks. Medical image analysis **68**, 101871 (2021)
21. Riesenburger, R.I., Safain, M.G., Ogbuji, R., Hayes, J., Hwang, S.W.: A novel classification system of lumbar disc degeneration. Journal of Clinical Neuroscience **22**(2), 346–351 (2015)
22. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence **1**(5), 206–215 (2019)

23. Shah, J., Siddiquee, M.M.R., Su, Y., Wu, T., Li, B.: Ordinal classification with distance regularization for robust brain age prediction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7882–7891 (2024)
24. Shen, L., Zheng, J., Lee, E.H., Shpanskaya, K., McKenna, E.S., Atluri, M.G., Plasto, D., Mitchell, C., Lai, L.M., Guimaraes, C.V., et al.: Attention-guided deep learning for gestational age prediction using fetal brain MRI. Scientific reports **12**(1), 1408 (2022)
25. Sneath, R.J., Khan, A., Hutchinson, C.: 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)
26. Teenbaum, J., Silva, D., Langford, J.: global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
27. Vaseli, H., Gu, A.N., Ahmadi Amiri, S.N., Tsang, M.Y., Fung, A., Kondori, N., Saadat, A., Abolmaesumi, P., Tsang, T.S.: Protoasnet: Dynamic prototypes for inherently interpretable and uncertainty-aware aortic stenosis classification in echocardiography. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 368–378. Springer (2023)
28. Yi, W., Zhao, J., Tang, W., Yin, H., Yu, L., Wang, Y., Tian, W.: 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)
29. Zheng, H.D., Sun, Y.L., Kong, D.W., Yin, M.C., Chen, J., Lin, Y.P., Ma, X.F., Wang, H.S., Yuan, G.J., Yao, M., et al.: Deep learning-based high-accuracy quantitation for lumbar intervertebral disc degeneration from MRI. Nature communications **13**(1), 841 (2022)
30. Zhou, L., Zhang, Y., Zhang, J., Qian, X., Gong, C., Sun, K., Ding, Z., Wang, X., Li, Z., Liu, Z., et al.: Prototype learning guided hybrid network for breast tumor segmentation in DCE-MRI. IEEE Transactions on Medical Imaging (2024)