Automatic grading of intervertebral disc degeneration in lumbar dog spines

Frank Niemeyer, Fabio Galbusera, Martijn Beukers, René Jonas, Youping Tao, Marion Fusellier, Marianna A Tryfonidou, Cornelia Neidlinger-wilke, Annette Kienle, Hans-joachim Wilke

Received: 4 August 2023 | Revised: 20 February 2024 | Accepted: 25 February 2024

DOI: 10.1002/jsp2.1326

RESEARCH ARTICLE

JOR Spine OPEN ACCESS

Automatic grading of intervertebral disc degeneration in lumbar dog spines

Frank Niemeyer[1,2] | Fabio Galbusera[1,2,3] | Martijn Beukers[4] | René Jonas[1] | Youping Tao[2] | Marion Fusellier[5] | Marianna A. Tryfonidou[4] | Cornelia Neidlinger-Wilke[1,2] | Annette Kienle[2] | Hans-Joachim Wilke[1,2]

[1]Institute for Orthopaedic Research and Biomechanics, Centre for Trauma Research, University Hospital Ulm, Ulm, Germany

[2]SpineServ GmbH & Co. KG, Ulm, Germany

[3]Head Research Group Spine, Spine Center, Schulthess Clinic, Zürich, Switzerland

[4]Department of Clinical Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands

[5]Maitre de Conférences Imagerie Médicale, INSERM UMRS1229, Regenerative Medicine and Skeleton RMeS Team STEP, School of Dental Surgery, Nantes, France

Correspondence
Hans-Joachim Wilke, Institute for Orthopaedic Research and Biomechanics, Centre for Trauma Research, University Hospital Ulm, Helmholtzstr. 14, 89081 Ulm, Germany.
Email: hans-joachim.wilke@uni-ulm.de

Funding information
European Union's Horizon 2020 research and innovation program iPSpine, Grant/Award Number: 825925

Abstract

Background: Intervertebral disc degeneration is frequent in dogs and can be associated with symptoms and functional impairments. The degree of disc degeneration can be assessed on T2-weighted MRI scans using the Pfirrmann classification scheme, which was developed for the human spine. However, it could also be used to quantify the effectiveness of disc regeneration therapies. We developed and tested a deep learning tool able to automatically score the degree of disc degeneration in dog spines, starting from an existing model designed to process images of human patients.

Methods: MRI midsagittal scans of 5991 lumbar discs of dog patients were collected and manually evaluated with the Pfirrmann scheme and a modified scheme with transitional grades. A deep learning model was trained to classify the disc images based on the two schemes and tested by comparing its performance with the model processing human images.

Results: The determination of the Pfirrmann grade showed sensitivities higher than 83% for all degeneration grades, except for grade 5, which is rare in dog spines, and high specificities. In comparison, the correspondent human model had slightly higher sensitivities, on average 90% versus 85% for the canine model. The modified scheme with the fractional grades did not show significant advantages with respect to the original Pfirrmann grades.

Conclusions: The novel tool was able to accurately and reliably score the severity of disc degeneration in dogs, although with a performance inferior than that of the

Frank Niemeyer and Fabio Galbusera are equal contributors to this work and designated as co-first authors.

human model. The tool has potential in the clinical management of disc degeneration in canine patients as well as in longitudinal studies evaluating regenerative therapies in dogs used as animal models of human disorders.

KEYWORDS
canine spine, deep learning, degeneration, image analysis, machine learning, radiological classification

1 | INTRODUCTION

Intervertebral disc (IVD) degeneration is a relatively common finding in dogs and are presented with variable spinal diseases covering predilection sections, dependent on the breed and the spinal disease[1] of the cervical, thoracolumbar, and lower lumbar spine. Disc degeneration can be painful and may lead to clinical disorders dependent on the underlying pathology and localization such as herniations and stenosis, but asymptomatic cases with concomitant disc degeneration are also not uncommon.[2,3] Treatments include rest, anti-inflammatory medication, and surgery dependent on the affected IVD segment such as discectomy, spinal canal decompression, and stabilization with plates and vertebral body screws, which showed good clinical results in selected cases.[4,5]

In recent years, a rising interest in grading schemes for the severity of IVD degeneration has been demonstrated by the publication of papers in which existing methods used for human patients have been tested in client-owned dogs,[2,6] papers describing novel schemes specific to dog subjects,[3,7,8] and paper commonly using grading of disc degeneration in experimental studies within the field of regenerative medicine. In particular, the Pfirrmann score used to classify T2-weighted midsagittal MRI scans of human lumbar IVDs[9] has been tested on thoracolumbar dog spines,[2] although employing an open permanent magnet with a low field strength of 0.2 T, in contrast to the original study that was conducted with 1.5 T scanners. Despite this limitation, the authors could prove the validity of the scheme, which showed good intraobserver and interobserver reliabilities (Cohen's κ scores of 0.93 and 0.81, respectively). The study involved 994 dog discs and confirmed that degeneration increases with aging and has a higher prevalence in client-owned dogs suffering from spinal disease. Recently this grading scheme was revisited with 1.5 T MRI images.[10]

We recently proposed a deep learning tool able to automatically grade human lumbar IVDs based on the Pfirrmann scheme,[11] which showed excellent accuracy and reproducibility exceeding those of both expert human observers and available machine learning-based implementations. This tool can in principle be extended to process images of dog spines after fine-tuning with appropriate imaging data. Such an extension would offer significant opportunities, both in terms of veterinary care and human biomedical research. Its use in the clinical veterinary practice would improve the diagnosis of IVD degeneration by reducing the risk of missed cases and the variability between observers.[11–13] Furthermore, the availability of this tool would potentially allow a more precise and objective grading of IVDs when monitoring the progression of the degeneration in longitudinal studies, or even the possible regression of the degenerative findings in the case of regenerative therapies. The latter case is especially relevant in light of the possible use of dogs as an animal model for human disc degeneration, which has been extensively reported in the literature.[14–17]

The aim of this paper is to develop an AI-based tool able to automatically apply the Pfirrmann scheme for IVD degeneration to T2-weighted midsagittal images of lumbar dog spines and to test the performance of the AI-based model against a set of images evaluated by two expert human observers. The secondary aims of the study are a critical assessment of the Pfirrmann scheme when used on dog discs instead of the human ones for which it was developed, and its possible extension to capture more subtle differences among the various grades.[11]

2 | MATERIALS AND METHODS

2.1 | Data collection

Client dog owners approved the use of the anonymized imaging data for research purposes. We collected MRI data of dog thoracolumbar spines of various breeds obtained in two veterinary clinics at Utrecht University (734 dogs) and the University of Nantes (101 dogs). The investigated subjects covered a wide age range and were subjected to MRI examinations for clinical reasons. For 340 of the subjects of the Utrecht dataset, information about breed, age, weight, and sex has been collected and is reported in detail in a previous study[18] (154 females (111 castrated) and 186 males (87 castrated), median age 6 years (range: 4 months-15.5 years), median weight 25.1 kg (range: 2.5–88 kg)). The subjects recruited in Nantes had median age of 6 years (range: 1–14) and median weight of 13.6 kg (range: 4.5–62.5 kg). No information was available for the remaining 394 dogs. Images were acquired with a Philips Ingenia 1.5 T scanner (Philips Healthcare, Eindhoven, The Netherlands) at the first recruiting site and with a Siemens Magnetom Essenza 1.5 T (Siemens AG, Erlangen, Germany) at the second one. Midsagittal T2-weighted images were considered for the evaluation of the degree of IVD degeneration. In both sites, images were acquired with a turbo spin echo sequence and slice thickness that was optimized for patient size (range: 2–2.5 mm). Higher acquisition times were used for smaller dogs with respect to the larger ones in order to achieve a sufficient signal-to-noise ratio.

2.2 | Data evaluation

The images were processed with purposely developed C++ software in order to select a region of interest covering the IVD and approximately half of the adjacent vertebral bodies. Only IVDs in the T13-S1 region that were clearly visible and with sufficient image quality were selected; this resulted in 5347 discs from the Utrecht recruitment site and 644 discs from the site in Nantes. These 5991 disc images were then merged into a single database.

Two expert human observers, an orthopedic surgeon (YT) and a veterinary diagnostic imaging specialist (MB), evaluated the degree of degeneration following a classification scheme based on Pfirrmann's algorithm.[9] The modified scheme was introduced by Niemeyer et al.[11] and included transitional grades (1.3, 1.7, 2.3, 2.7, 3.3, 3.7, 4.3, 4.7) which aim at expressing uncertainties and tendencies of discs that do not fit exactly the algorithm described in the original classification (Figure 1, Table 1). For example, the grade 2.3 indicates a disc that shows more evident signs of degeneration than grade 2, for example, a slightly unclear distinction between the nucleus pulposus and the annulus fibrosus, with the IVD however fitting more closely to grade 2 than grade 3.

The two human raters (MB, YT) used the modified scheme to classify all discs in the dataset; besides, one of the two raters examined the whole dataset twice in order to allow for the calculation of the intraobserver agreement by means of Cohen's κ. For the first 801 discs, the two raters discussed the differences between the evaluations and, in case of disagreement, reached a unique conclusion through personal discussion. The remaining 5491 discs were reevaluated by one of the two observers only, who took advantage of the knowledge gained in the consensus phase. The interobserver agreement was then calculated by means of Cohen's κ statistics by comparing the evaluations of the two raters for the first 801 discs before the consensus meeting.

Since the modified classification scheme can be directly converted into the equivalent Pfirrmann grades (Table 1), the analysis of the intraobserver and interobserver reliability was also conducted for the original Pfirrmann scheme after pooling the appropriate grades together.

2.3 | Pre-processing

After adjusting the aspect ratio of the regions of interest to 1:1, square images containing individual IVDs were cropped from the

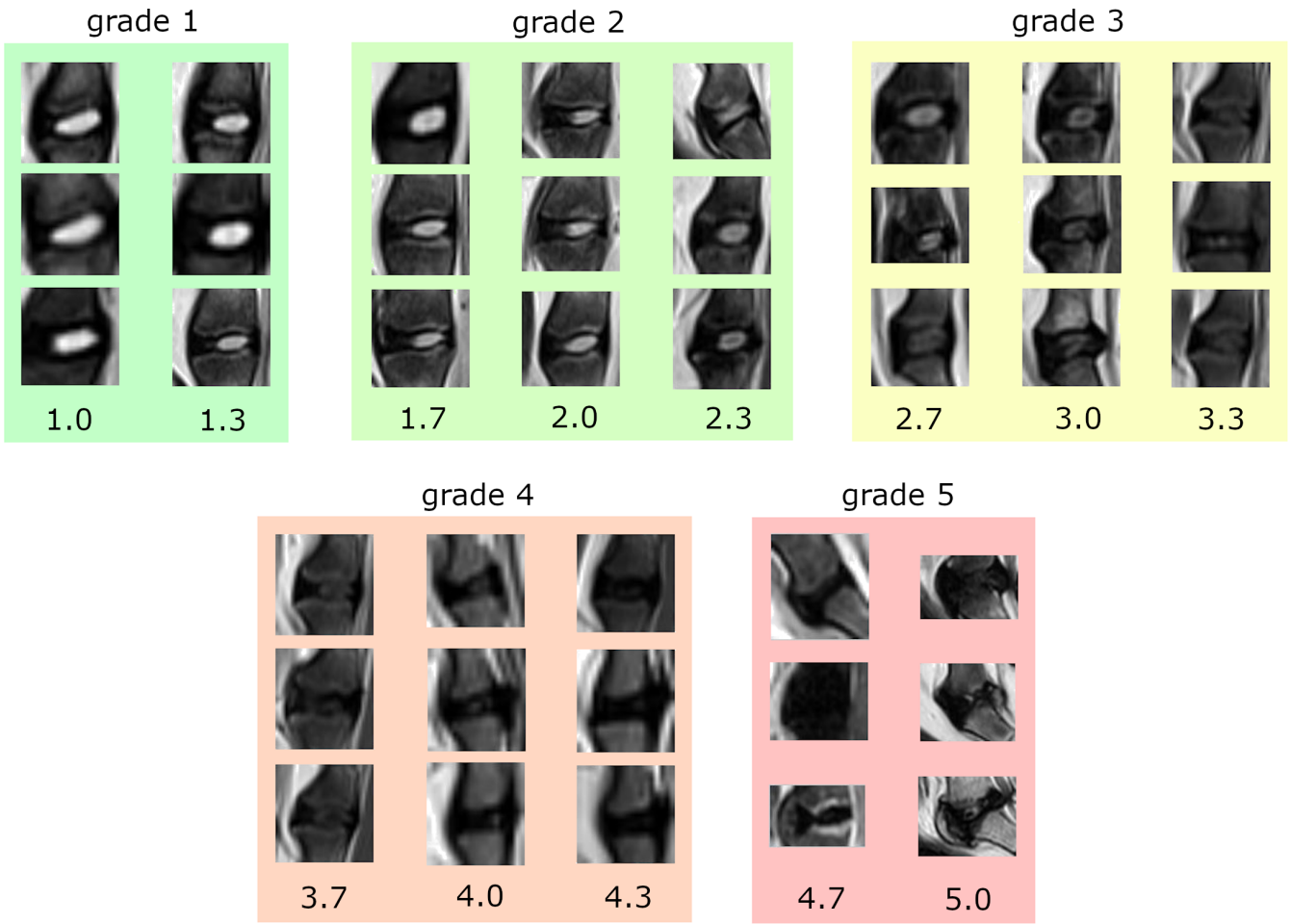

FIGURE 1 Examples of canine lumbar discs belonging to the various grades of the Pfirrmann scale[9] as well as to the modified one including fractional grades.

TABLE 1 Definitions of the various grades of the modified Pfirrmann classification scheme including fractional grades.[11]

Grade	Equivalent Pfirrmann grade	Description
1.0	1	Homogeneous bright white structure
1.3	1	Bright white structure with minor signs of inhomogeneity
1.7	2	Bright white structure with local inhomogeneity
2.0	2	Inhomogeneous white structure, possible horizontal bands
2.3	2	White structure with local loss of signal, possible horizontal bands
2.7	3	Inhomogeneous gray structure with bright regions, clear distinction between nucleus and annulus
3.0	3	Clear loss of signal, clear distinction between nucleus and annulus
3.3	3	Clear loss of signal, visible distinction between nucleus and annulus
3.7	4	Barely visible distinction between nucleus and annulus, preserved disc height
4.0	4	No distinction between nucleus and annulus, preserved disc height
4.3	4	No distinction between nucleus and annulus, minor height loss
4.7	5	Partially collapsed disc space
5.0	5	Collapsed disc space

original images. The resulting images were then resized to 128 × 128 and normalized to [−1, 1], preserving the original 16-bit depth of the MRI scans. In order to improve the robustness of the predictions, data augmentation was implemented by randomly flipping the images horizontally, performing a random rotation of ±20 degrees, randomly resizing the region of interest (before cropping) to 80%–100% of its original height or width and randomly shifting the region of interest in both horizontal and vertical direction by up to ±10% of its width. This way, for each of the collected IVD segment images, seven additional images were generated by and added to the training set (i.e., eight-fold augmentation).

2.4 | Neural network architecture, losses, and training

The architecture of the neural network used to perform the evaluations was based on the one presented in Lee et al.,[11] with minor changes aimed at minimizing overfitting. The model was designed to use a T2-weighted image of dog disc as input, and to provide the degeneration grade based on the modified classification scheme as output; the equivalent Pfirrmann grade could then be derived by pooling the appropriate grades (Table 1). In brief, the classifier was based on the VGG-16 network[19] with weights initialized as done in a previous study.[20] With respect to VGG-16, ReLU activations were replaced by leaky ReLUs, post-activation batch normalization was added after each convolutional layer, and the number of max pooling layers was reduced by using dilated convolutions. The top fully connected layers were replaced by a global average pooling layer, significantly reducing the number of parameters and therefore overfitting as well as the memory footprint. With respect to the implementation in Lee et al.,[11] dropout and L2 regularization were used more aggressively in order to further prevent overfitting. The classifier was implemented using the TensorFlow 2.1.0 library (Google LLC, Mountain View, CA, USA), and trained on a workstation equipped with an NVIDIA Titan RTX card (NVIDIA Corp., Santa Clara, CA, USA). Similar to the previous study, the available data was randomly split so that 90% was used for training and hyperparameter tuning, whereas the remaining 10% was used to test the performance of the model.

2.5 | Data evaluation

The metrics used for the evaluation of the performance of the tool were: (1) accuracy, that is the exact agreement between ground truth and prediction; (2) sensitivity; (3) specificity; (4) Matthews correlation coefficients (MCC). The assessment was performed on a grade-specific level, that is, considering the ability of the neural networkin predicting each individual grade, as well as the average for all grades calculated by summing the numbers of true positives, true negatives, false positives, and false negatives of the individual grades. The values of the metrics were compared with those calculated for human IVDs with the default model described in Lee et al.,[11] using the same dataset as the original publication.

3 | RESULTS

The analysis of the intraobserver and interobserver agreements conducted on the evaluations before the consensus meeting showed a moderate agreement for the modified classification scheme (Cohen's κ equal to 0.43 in both cases), whereas the agreement was substantial for the original Pfirrmann scheme (Cohen's κ equal to 0.67) (Table 2).

The prevalence of the various grades of IVD degeneration as manually assessed by the human observers, which included a specialist in veterinary radiology, shows a different distribution in the dog population with respect to those described in human subjects suffering from low back pain (Figure 2). Dog spines showed a markedly higher occurrence of non-degenerated discs, that is, grade 1, whereas in human patients grade-3 and grade-4 discs constituted the majority of the samples. Conversely, grade-5 discs were almost absent in the dog population, showing that the complete collapse of the IVD is a rare finding in dogs. In both dogs and humans, when the fractional grades (1.3, 1.7, etc.) were considered, they were employed less frequently by the evaluators than the original grades (1.0, 2.0, etc.),

TABLE 2 Analysis of the intraobserver and interobserver agreement before the consensus meeting.

Comparison	Cohen's *k*	exact	1 grade	2 grades	3 grades	4 grades	5 grades	6 grades	7 grades
Intraobserver agreement									
Pfirrmann grading (1–5)	0.67	74.6%	25.0%	0.4%	0%	0%	-	-	-
Modified (1.0, 1.3, etc.)	0.43	49.4%	31.8%	12.4%	5.3%	0.9%	0.1%	0%	0%
Interobserver agreement									
Pfirrmann grading (1–5)	0.67	75.4%	24.2%	0.3%	0%	0%	-	-	-
Modified (1.0, 1.3, etc.)	0.43	46.8%	35.3%	13.3%	4.0%	0.3%	0.1%	0.1%	0%

Note: "exact": exact agreement between the two evaluations; "1 grade": disagreement of 1 grade (e.g., 2 vs. 3 for the Pfirrmann grading, 3.7 vs. 4.0 for the modified scheme, etc.).

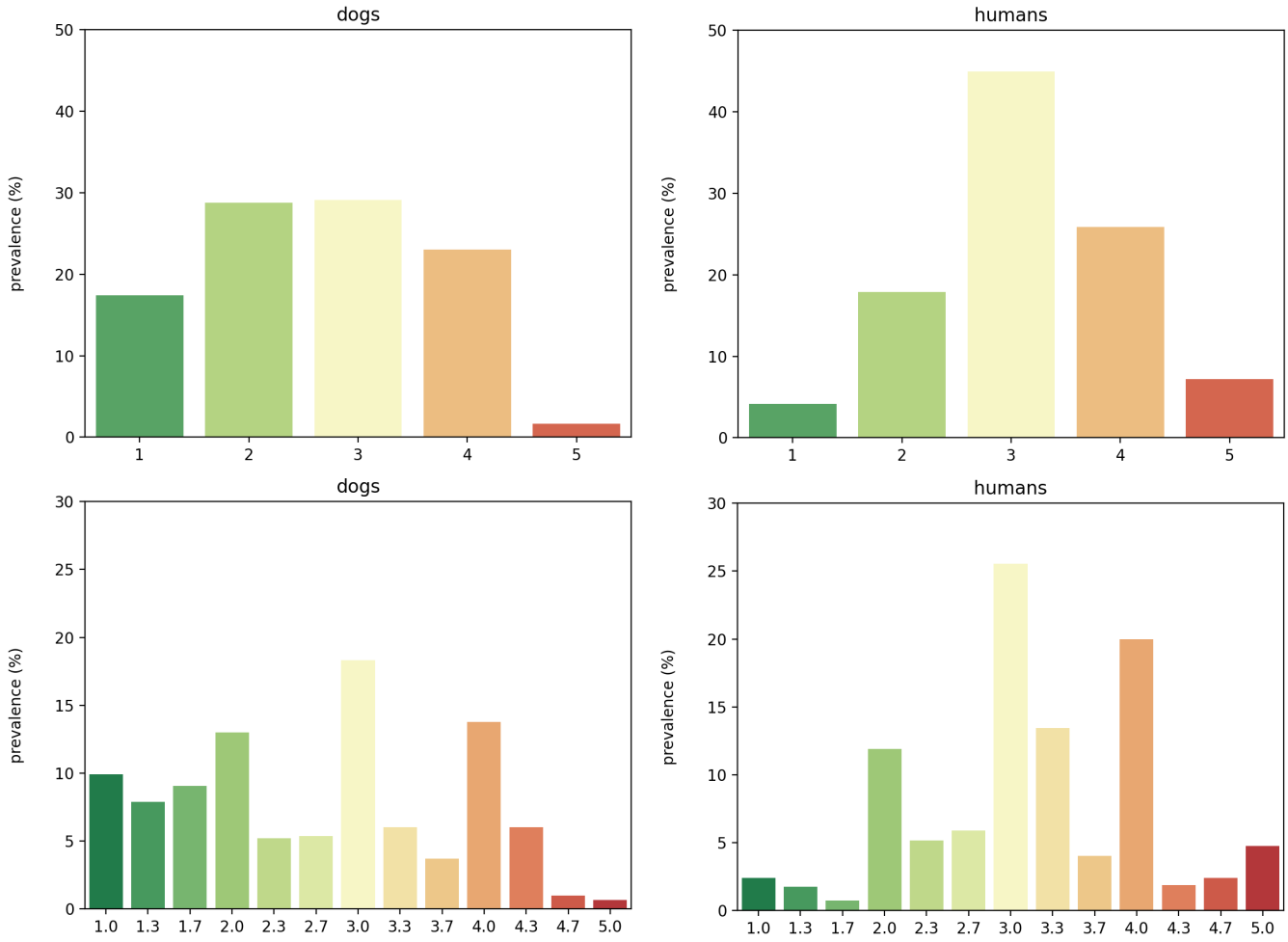

FIGURE 2 Prevalence of the individual grades of degeneration in canine spines (left) and in humans suffering from low back pain (right),[11] either based on the Pfirrmann classification scheme[9] (first row) and on the modified scheme including fractional grades (second row).

showing that the majority of discs fit well into the original definitions provided by Pfirrmann et al.[9]

In general, the prediction of the grade of disc degeneration of dog spines based on the Pfirrmann scheme showed a good performance (Table 3, Figure 3). Considering the average metrics among all grades, the model scored a sensitivity of 85.2% and a specificity of 96.3%. The worst performance was observed for grade 5 which showed a sensitivity of 16.3%, due to a number of grade-4 discs classified as grade-5 by the model and arguably to the rarity of grade-5 discs in the training data. Despite the generally good values, the metrics described a worse performance with respect to the model processing human images, in particular, lower sensitivities also for grades 2, 3, and 4 which were reflected by an evidently higher dispersion in the scatter plots (Figure 3).

The quality of the predictions decreased significantly when the modified scheme with the transitional grades was considered (Table 4,

Figure 4). Although the accuracies were consistently high due to the dominance of true negatives, sensitivities and specificities showed a wide range of variation, with low peaks for the fractional grades whereas the performance for the integral grades (1.0, 2.0, etc.) was generally better. The worst metrics were found for grade 4.7, for which no true positives were calculated. The model trained on human discs showed a better performance, with low sensitivities for some fractional grades (1.3, 1.7, 3.7, 4.3, and 4.7) which were relatively weakly represented in the training dataset.

The analysis of the individual dog discs with the largest difference between ground truth and predictions highlighted some interesting patterns (Figure 5). One disc showing collapse of the IVD space, thus labeled as grade 5, also exhibited a high signal in the nucleus pulposus which arguably triggered the grade-1 prediction; this sample indeed did not fit the Pfirrmann scheme since features describing both low (high signal) and severe degeneration (disc space collapse) were simultaneously present. Other interesting discs showed hyperintensity in the vertebra due to inflammation which was arguably mistaken as the disc, determining a better score with respect to the ground truth.

TABLE 3 Accuracy, sensitivity, specificity, and Matthews correlation coefficient (MCC) calculated for the classifiers processing canine and human images,[11] for the Pfirrmann score of intervertebral disc degeneration.[9]

Dogs				
Grade	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**MCC**
1	95.5	91.5	96.4	0.85
2	91.3	84.6	94.0	0.78
3	91.0	83.2	94.2	0.78
4	93.9	88.5	95.5	0.83
5	98.6	16.3	99.9	0.37
Average	94.1	85.2	96.3	0.81
Humans				
Grade	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**MCC**
1	99.1	89.3	99.6	0.89
2	95.1	93.6	98.3	0.91
3	96.2	95.9	93.7	0.90
4	98.1	89.6	98.4	0.90
5	98.1	81.6	99.4	0.85
Average	97.6	89.4	97.9	0.90

4 | DISCUSSION

This paper describes the development and testing of a deep learning tool able to classify the degree of disc degeneration in midsagittal MRI scans of dog lumbar IVDs, using an existing AI-based model able to process human images as a starting point. In general, the novel tool showed a performance sufficient to warrant its use in the clinical practice as well as in basic research studies in which dogs are used as an animal model for the human pathology to study the potential of

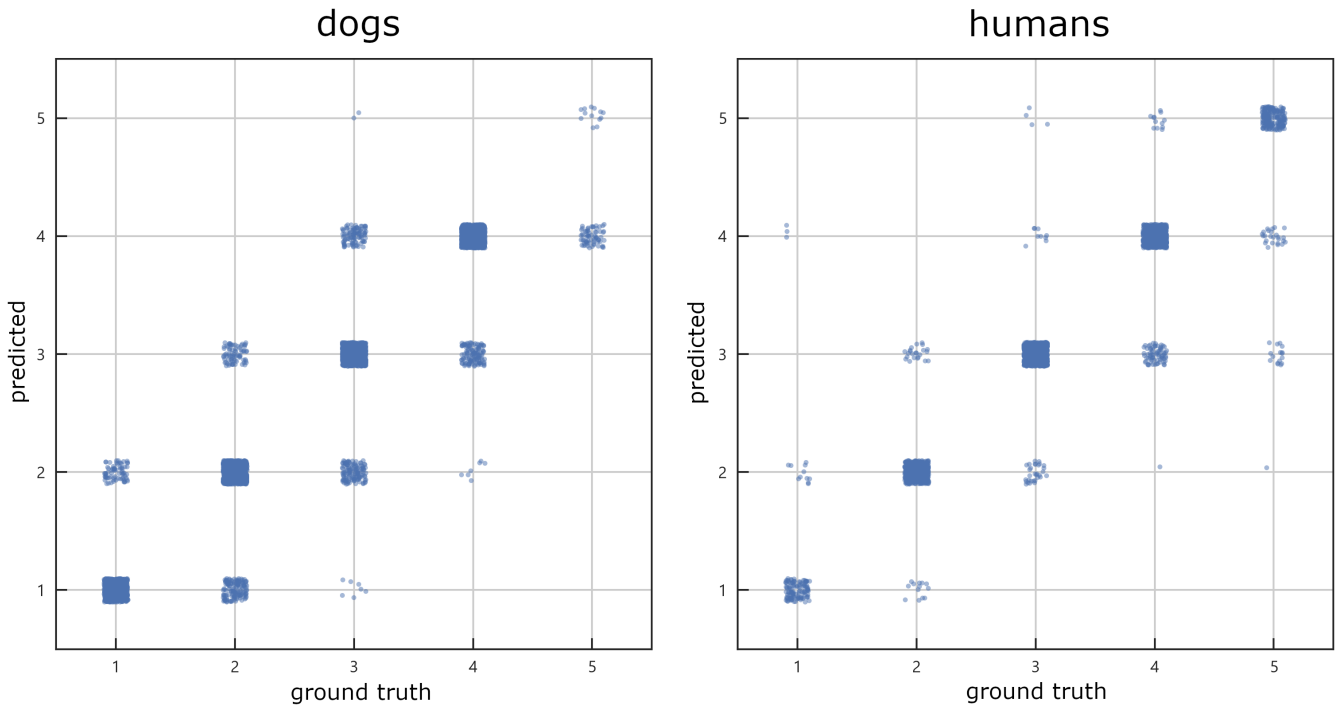

FIGURE 3 Ground truth versus grades predicted by the models on canine spines (left) and discs of human patients suffering from low back pain (right), based on the Pfirrmann classification scheme.[9] All points have been jittered randomly by ±0.1 Pfirrmann grades to improve the visibility of individual samples.

TABLE 4 Accuracy, sensitivity, specificity and Matthews correlation coefficient (MCC) calculated for the classifiers processing canine and human images,[11] for the modified Pfirrmann score with fractional grades.

Dogs				
Grade	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**MCC**
1.0	96.1	82.0	97.6	0.78
1.3	93.0	68.9	95.1	0.57
1.7	92.8	38.9	98.1	0.48
2.0	91.9	92.6	93.3	0.69
2.3	93.6	33.5	96.9	0.32
2.7	94.4	30.5	98.0	0.35
3.0	90.6	83.5	92.2	0.70
3.3	93.7	28.8	97.9	0.34
3.7	95.4	17.0	98.4	0.20
4.0	92.4	85.1	93.6	0.72
4.3	97.7	69.4	99.5	0.78
4.7	99.0	0.0	100.	-
5.0	99.6	46.9	99.9	0.64
Average	94.6	65.1	97.1	0.62
Humans				
Grade	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**	**MCC**
1.0	98.5	96.1	98.6	0.77
1.3	98.2	10.7	99.7	0.21
1.7	99.2	12.5	99.8	0.20
2.0	97.0	89.2	98.0	0.86
2.3	97.2	73.1	98.3	0.70
2.7	97.6	69.7	99.6	0.78
3.0	95.8	95.3	96.0	0.89
3.3	95.6	88.0	96.7	0.82
3.7	96.1	29.7	98.9	0.38
4.0	95.7	92.9	96.4	0.87
4.3	98.6	33.3	99.9	0.53
4.7	98.3	36.8	99.8	0.54
5.0	97.9	89.5	98.3	0.80
Average	97.3	82.8	98.5	0.81

regenerative therapies. However, the results clearly indicated a performance gap with respect to those referring to images of the human spine.

Such a finding may be attributed to several reasons. First, the Pfirrmann scheme has been constructed based on the radiological findings typically observed on images of human patients, and not those on dog images. Radiological features that are common in humans involved in the degenerative cascade may be rare in dogs,[2] for example, the collapse of the IVD space which was observed only in 2% of the dog images. Besides, dog discs depart from a grade I characteristic of a much higher hydration level based on the T2 signal evident in the high number of grade-1 discs in the present data set, while the majority of human discs typically exhibit a loss of T2 signal, the first visible degenerative sign with relatively little to no discs being graded as grade-1. Indeed, studies showed that although IVD degeneration shares many features among humans and dogs, the biomechanics and biology of the degenerative cascade show remarkable differences between the two species and even among canine breeds.[1,15,21,22] Furthermore, human IVDs are significantly larger and therefore offer significantly better visualization of the degenerative features with MRI techniques and scanners which have been indeed developed to match the size and properties of the human body. Subtle degenerative changes, which would be evident in images of human patients, may simply not be observable in dog MR scans due to the lower resolution of the technique with respect to the size of the anatomical structure of interest—using standard-of-care imaging settings. It should also be noted that the deep learning model processing human images has been developed through multiple iterations in which the training dataset has been optimized by adding images of the least represented degeneration grades, whereas this procedure was not conducted in the present study due to the relatively lower availability of high-quality images of dog subjects and the underrepresentation of grade-5 discs.

To our knowledge, this study is the first one presenting an automated tool aimed at grading the degeneration of dog discs, and direct comparisons with the existing literature are therefore not possible. However, it should be noted that any evaluation by human observers, such as the one presented in Bergknut et al.,[2] has a limited reproducibility whereas machine learning-based tools do not suffer from such an issue. This advantage may be critical in applications in which capturing subtle changes over time of the same disc is important, such as in pre-clinical studies evaluating regenerative therapies for degenerative disc disease with longitudinal imaging investigations.[17] A valid alternative approach to machine learning-based tools is offered by quantitative methods such as T2 mapping,[23] which directly provide a reproducible numerical outcome based on the physics of the structure under investigation. However, quantitative imaging requires special equipment and/or software which may not be available in all clinical settings, especially in veterinary clinics.

In general, the modified scheme with fractional grades did not show significant advantages with respect to the original Pfirrmann classification system. As a matter of fact, the human observers tended to use the additional grades less frequently than the original ones (Figure 2), resulting in a degraded performance of the model. The same observation may apply also to human discs; in both cases, the grades defined in the original classification scheme seem to cover most of the imaging appearances of degenerated discs. However, some specific degrees may be considered as an exception since they indeed describe degeneration patterns commonly observed in the population. An example is grade 4.7 which was found in more than 5% of the dog population, in which major disc collapse is rare but a minor height loss is relatively common in severely degenerated T2 hypointense (black) discs. In this case, the deep learning model showed a relatively good performance with a sensitivity of 69.4%, a specificity of 99.5%, and an MCC of 0.78.

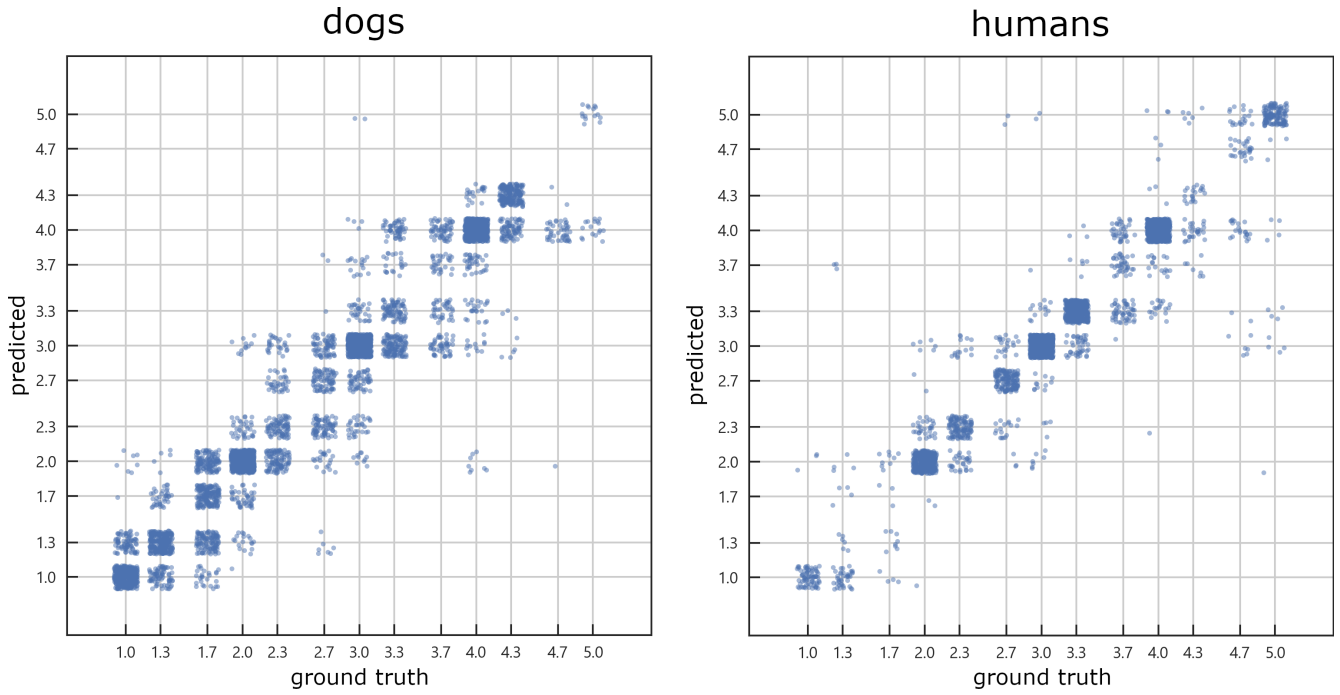

FIGURE 4 Ground truth versus grades predicted by the models on canine spines (left) and discs of human patients suffering from low back pain (right), based on the modified classification scheme including fractional grades. All points have been jittered randomly by ±0.1 Pfirrmann grades to improve the visibility of individual samples.

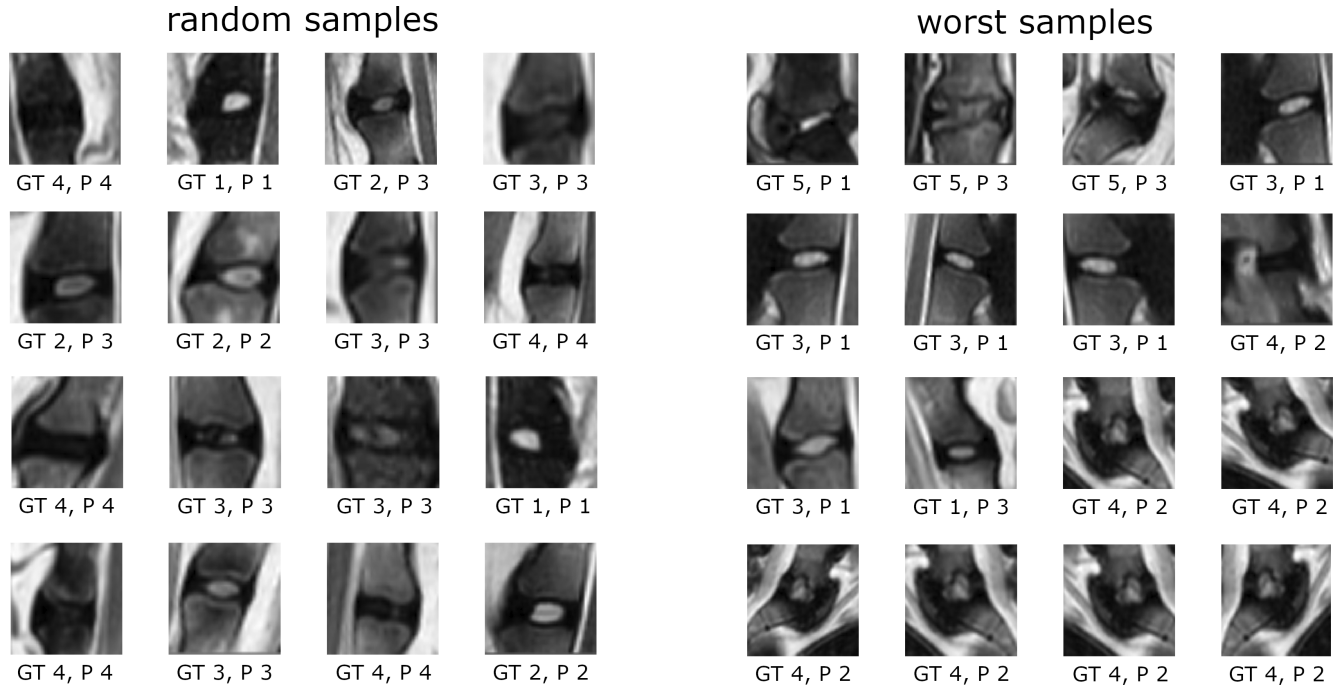

FIGURE 5 Randomly selected samples (left) and those showing the largest difference between ground truth and predicted grade (right). "GT": ground truth; "P": predicted grade.

As with all studies, the present investigation has limitations. The training and test sets were relatively small and included images obtained with 1.5 T MRI scanners, which offer substantially higher image quality with respect to the equipment available in many veterinary clinical settings.[24] Since images acquired with a lower magnetic field have reduced resolution and information content, an inferior performance of the model should be expected; the same would apply, however, to human observers. Another limitation pertains to

the fact that the ground truth was determined by a single rater, whereas the consensus agreement was conducted only on less than 10% of the available data.

In conclusion, the deep learning tool presented in this study demonstrated to be able to accurately and reliably score the disc degeneration degree of dog spines, although with an average performance inferior to the original model processing images of the human spine from which the present tool is derived. The novel tool opens new perspectives in the clinical management of disc degeneration in dog patients and may provide critical advantages in settings taking advantage of a reproducibility superior to that of human observers, such as for example in longitudinal studies evaluating regenerative therapies in dogs intended as a model as the human pathology adding an objective means of evaluation to the toolbox of experimental studies in disc degeneration/regeneration.[10]

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program iPSpine under grant agreement No. 825925. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Marion Fusellier https://orcid.org/0000-0001-5336-9352
Marianna A. Tryfonidou https://orcid.org/0000-0002-2333-7162
Hans-Joachim Wilke https://orcid.org/0000-0001-6007-8844

REFERENCES

1. Smolders LA, Bergknut N, Grinwis GCM, et al. Intervertebral disc degeneration in the dog. Part 2: Chondrodystrophic and non-chondrodystrophic breeds. *Vet J*. 2013;195(3):292-299.
2. Bergknut N, Auriemma E, Wijsman S, et al. Evaluation of intervertebral disk degeneration in chondrodystrophic and nonchondrodystrophic dogs by use of Pfirrmann grading of images obtained with low-field magnetic resonance imaging. *Am J Vet Res*. 2011;72(7):893-898.
3. Bergknut N, Meij BP, Hagman R, et al. Intervertebral disc disease in dogs—Part 1: a new histological grading scheme for classification of intervertebral disc degeneration in dogs. *Vet J*. 2013;195(2):156-163.
4. Jeffery ND, Harcourt-Brown TR, Barker AK, Levine JM. Choices and decisions in decompressive surgery for thoracolumbar intervertebral disk herniation. *Vet Clin North Am Small Anim Pract*. 2018;48(1):169-186.
5. Worth A, Fitzpatrick N, Costa-Valente R, et al. Canine degenerative lumbosacral stenosis: prevalence, impact, and management strategies. *Vet Med*. 2019;10:169-183. doi:10.2147/VMRR.S180448
6. Bergknut N, Grinwis G, Pickee E, et al. Reliability of macroscopic grading of intervertebral disk degeneration in dogs by use of the Thompson system and comparison with low-field magnetic resonance imaging findings. *Am J Vet Res*. 2011;72(7):899-904.
7. Besalti O, Pekcan Z, Sirin YS, Erbas G. Magnetic resonance imaging findings in dogs with thoracolumbar intervertebral disk disease: 69 cases (1997-2005). *J Am Vet Med Assoc*. 2006;228(6):902-908.
8. Downes CJ, Gemmill TJ, Gibbons SE, McKee WM. Hemilaminectomy and vertebral stabilisation for the treatment of thoracolumbar disc protrusion in 28 dogs. *J Small Anim Pract*. 2009;50(10):525-535.
9. Pfirrmann CW, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine*. 2001;26(17):1873-1878.
10. Lee NN, Salzer E, Bach FC, Cook JL. A comprehensive tool box for large animal studies of intervertebral disc degeneration. *JOR Spine*. 2021;4(2):e1162.
11. Niemeyer F, Galbusera F, Tao Y, Kienle A, Beer M, Wilke HJ. A deep learning model for the accurate and reliable classification of disc degeneration based on MRI data. *Invest Radiol*. 2021;56(2):78-85. doi:10.1097/RLI.0000000000000709
12. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine*. 2019;2(1):e1044.
13. Jamaludin A, Lootus M, Kadir T, et al. ISSLS prize in bioengineering science 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J*. 2017;26(5):1374-1383.
14. Alini M, Eisenstein SM, Ito K, et al. Are animal models useful for studying human disc disorders/degeneration? *Eur Spine J*. 2008;17(1):2-19.
15. Bergknut N, Rutges JPHJ, Kranenburg HJC, et al. The dog as an animal model for intervertebral disc degeneration? *Spine*. 2012;37(5):351-358.
16. Lee NN, Kramer JS, Stoker AM, et al. Canine models of spine disorders. *JOR Spine*. 2020;3(4):e1109.
17. Mern DS, Walsen T, Beierfuß A, Thomé C. Animal models of regenerative medicine for biological treatment approaches of degenerative disc diseases. *Exp Biol Med*. 2021;246(4):483-512.
18. Beukers M, Grinwis GCM, Vernooij JCM, et al. Epidemiology of Modic changes in dogs: prevalence, possible risk factors, and association with spinal phenotypes. *JOR Spine*. 2023;6(3):e1273.
19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014 arXiv [cs.CV]. arXiv. http://arxiv.org/abs/1409.1556
20. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. *Proceedings of the IEEE International Conference on Computer Vision*. IEEE Xplore; 2015:1026-1034. openaccess.thecvf.com
21. Goggin JE, Li AS, Franti CE. Canine intervertebral disk disease: characterization by age, sex, breed, and anatomic site of involvement. *Am J Vet Res*. 1970;31(9):1687-1692.
22. Lotz JC. Animal models of intervertebral disc degeneration: lessons learned. *Spine*. 2004;29(23):2742-2750.
23. Chen C, Jia Z, Han Z, et al. Quantitative T2 relaxation time and magnetic transfer ratio predict endplate biochemical content of intervertebral disc degeneration in a canine model. *BMC Musculoskelet Disord*. 2015;16:157.
24. Farrelly J, McEntee MC. A survey of veterinary radiation facilities in 2010. *Vet Radiol Ultrasound*. 2014;55(6):638-643.

How to cite this article: Niemeyer F, Galbusera F, Beukers M, et al. Automatic grading of intervertebral disc degeneration in lumbar dog spines. *JOR Spine*. 2024;7(2):e1326. doi:10.1002/jsp2.1326