A Deep Learning Method for Identifying Predictors of Knee Osteoarthritis Radiographic Progression From Baseline MRI

Jean-Baptiste Schiratti
owkin
Rémy Dubois
Owkin
Paul Herent
Owkin
David Cahané
Owkin
Jocelyn Dachary
Owkin
Thomas Clozel
Owkin
Gilles Wainrib
Owkin
Florence Keime-Guibert
Laboratoires Servier: Les Laboratoires Servier SAS
Agnes Lalande
Laboratoires Servier: Les Laboratoires Servier SAS
Maria Pueyo
Laboratoires Servier: Les Laboratoires Servier SAS
Romain Guillier
Laboratoires Servier: Les Laboratoires Servier SAS
Christine Gabarroca
Laboratoires Servier: Les Laboratoires Servier SAS
Philippe Moingeon (philippe.moingeon@servier.com)
Laboratoires Servier: Les Laboratoires Servier SAS https://orcid.org/0000-0002-2380-9983

Research article

Keywords: knee osteoarthritis, Osteoarthritis Initiative database, classification model

Posted Date: May 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-521841/v1

Abstract

– Background –

The identification of patients with knee osteoarthritis (OA) likely to progress rapidly in terms of structure is critical to facilitate the development of disease-modifying drugs.

– Methods –

Using data from the Osteoarthritis Initiative database (OAI), we implemented a Deep Learning method to predict, from baseline magnetic resonance images, further cartilage degradation, the latter being measured by Joint Space Narrowing at 12 months.

– Results –

Using COR IW TSE images, our classification model achieved a ROC AUC score of 65% to be compared with a ROC AUC score of 58.7% obtained by trained radiologists. Additional analyses conducted in parallel to predict pain grade evaluated by the WOMAC pain index achieved a ROC AUC score of 72%. Attention maps provided evidence for distinct specific areas as being relevant in those two predictive models, including the internal femoro-tibial compartment for JSN progression and the intra-articular space for pain prediction.

– Conclusions –

This feasibility study demonstrates the interest of deep learning applied to OA, with a potential to support even trained radiologists in the challenging task of identifying patients with a high-risk of disease progression.

1. Introduction

Osteoarthritis (OA) is a common disease which constitutes the fourth leading cause of disability worldwide [1]. According to the US National Health Interview Survey, up to 14 million American people are considered to have a symptomatic knee [2], with additional tens of millions affected as well in Europe, South America, Asia, or Middle East [3]. As a consequence of ensuing healthcare expenditures and losses of activity, the economic burden associated with OA is estimated to represent up to 2.5% of Growth National Product in Western countries [4].

The standard of care for OA based on both non-pharmacological and symptomatic pharmacological treatments has only a limited effect on function and pain. Thus, a very high unmet medical need still persists for a disease-modifying osteoarthritis drug (DMOAD) counteracting disease progression for both function and pain and avoiding the requirement for knee surgical replacement. As of today, the development of such DMOADs has been unsuccessful for two reasons. First of all, significant differences are observed among patients in terms of progression of cartilage degradation. Secondly, in the absence

of any established patient stratification in the form of endotypes reflecting well-characterized pathophysiological mechanisms, the slow and heterogeneous evolution of the disease makes it difficult to evaluate the effectiveness of a treatment in a broad patient population, within the 1 or 2 year(s) usual timeframe of a clinical study [5].

In this context, a personalized medicine approach is being considered to treat OA, consisting in identifying the most appropriate target populations predicted to benefit from DMOADs [6]. Primary efficacy endpoints required to document DMOAD efficacy include both clinical variables such as requirement for joint replacement as well as structural changes. The diagnosis of knee OA and the evaluation of its severity are currently based on imaging, with radiography remaining the most commonly used modality in clinical practice [7]. Specifically, knee X-rays are used to determine the JSW (*Joint Space Width*) as a measurement of the distance between tibia and femur considered as an indicator of cartilage thickness. X-rays of the knee performed for an individual patient at various time points allow to define the JSN (*Joint Space Narrowing*) as a change in JSW over time [8]. Current regulatory guidelines for clinical trials aiming at evaluating candidate DMOADs recommend that JSN should be used as the primary endpoint in those trials [9].

One limitation, however, is that a reliable evaluation of JSN during patient follow-up remains difficult [10]. A clustering method on OAI data during a 8 year follow-up concluded that only 29% of patients displayed a radiographic progression (as defined by JSN), with no further association between progression and pain worsening [11]. In this context, the use of MRI emerges as a better quantitative endpoint recommended for assessing morphological changes in knee cartilage during OA [12]. MRI allows the assessment of meniscal lesions such as root meniscal tears and extrusions known to be associated with OA progression [13, 14]. It also detects other lesions predictive of pain, such as the presence of synovitis and synovial fluid effusion [15] or bone marrow lesions [16].

We thus undertook the present feasibility study in support of the development of candidate DMOADs with the assumption that the latter should preferably be evaluated in patients likely to progress rapidly. To assess whether knee MR images collected at baseline could predict further cartilage degradation, we implemented a deep learning method using baseline MR images to build up a predictive model for future progression of knee OA, the latter being measured by JSN at 12 months. Additional analyses were conducted in parallel to predict pain grade evaluated by the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).

2. Methods

2.1 Description of the OAI database

The OsteoArthritis Initiative database is a public multi-center longitudinal database assembled by a consortium led by the National Institutes of Health in the US to help better understand and prevent the progression of knee OA [17]. At baseline, a total of 4796 patients had a bilateral standing knee radiograph

(X-Ray) and 3D knee MRI. Follow-up visits were done at 12, 24, 36, 48, 72 and 96 months (with 65% of patients enrolled at baseline having a 96-month follow-up visit). The knee MRI sequences include sagittal 3D DESS, coronal 2D IW TSE and sagittal 2D IW TSE fat-suppressed. A detailed description of MRI sequences from the OAI database can be found in Peterfy et al [18]. The database further contains clinical informations (age, sex, Body Mass Index [BMI]...), including as well results of pain assessment from WOMAC, a self-administered questionnaire encompassing for each visit up to 24 items divided into 3 subscales (*i.e.* pain, stiffness and physical function). Assessments from X-Rays such as Kellgren & Lawrence (KL) grade [19] and JSW were performed as well in the cohort at several locations in the medial and lateral compartments.

2.2 Endpoints used in the study

2.2.1 Joint Space Narrowing

OA progression, defined as cartilage degradation over time, was measured by using X-Ray images as the minimum JSW in the medial compartment of the knee. This semi-automated measurement was obtained at several time points (baseline, 12 months, 24 months). As proposed by Bruyere et al. [20], a 12 month-OA progressor was defined as a patient's knee exhibiting a JSN at 12 months lower than − 0.5mm: JSN(12 months) = JSW(12 months) - JSW(baseline) $\leq$ -0.5mm. The threshold of -0.5mm for minimum radiographic JSN was identified as clinically relevant in several studies, see Reginster et al [21] for a review. Since the JSN criteria was evaluated separately for each knee, a patient could be a 12 month-OA progressor for a single knee or both. Using this JSW variation as a threshold, we proceeded to identify from knee MRIs those patients predicted to lose at least 0.5mm of knee cartilage.

2.2.2 WOMAC pain score

A secondary objective was to study the prediction of pain encoded by the WOMAC score, using contemporary MR images and clinical data (see description of clinical variables in Table 1 in additional materials). Hence, this objective was not to build a model predictive of future evolutions of the disease, but rather to explain the current state of the disease, still exploiting a combination of imaging and clinical information.

2.2.3 Evaluation

The performance of models described below was evaluated using a five-fold cross-validation scheme and measured with the following metrics: area under the Receiver Operating Characteristic (ROC AUC score), area under the Precision-Recall (PR) curve and F1 score. These metrics are well suited to binary classification tasks which suffer from class imbalance.

2.3 Preprocessing of MRI data

Prior to feeding images into the model, several preprocessing steps have been applied sequentially in order to normalize the dataset, as illustrated in Fig. 1 and summarized below.

2.3.1 Image conversion

The OAI database exposes images in the DICOM format. In order to ease image reading and writing operations, each acquisition was converted to the NIFTI format (representing a full, three-dimensional image) by using the dcm2niix software [22].

2.3.2 Image orientation

The OAI database contains images of both knees for each patient. Specifically, left knee images are RAS (Right, Anterior, Superior)-oriented, while right knee images are LAS (Left, Anterior, Superior)-oriented. In order to homogenize the dataset, orientations were normalized for all images. To this aim, images of right knees have been “mirrored” along the sagittal-axial plane in order to look similar to images of left knees. This operation was performed using the NiPype python library [23].

2.3.4 Bias field correction

MR images can suffer from local magnetic field variations, resulting in artefacts in the reconstructed image. To solve this problem, the N4 bias field correction method [24] was applied to reconstructed images.

The raw MR image is first re-oriented so that both left and right knees are similarly oriented. Noteworthy, only the left knee image is flipped, whereas the right is maintained as is, in order to obtain uniform orientations across the dataset. The N4 bias field correction is then applied, together with a color normalization step.

2.4 Model architectures

2.4.1 Feature extraction

For both architectures used for 2D and 3D MR images, meaningful features were first extracted by using models trained on the ImageNet dataset [25], resulting in high performances on the associated challenge. This approach, most suitable to address problems of very high dimensionality, further allows to speed up model training by delegating the computationally cumbersome task of building meaningful representations from images, before feeding those representations into a classification neural network.

When applied to MR images, such representations are extracted from each individual slice of the image, resulting in up to 150 (respectively 30) 1280-dimensional feature vectors for SAG 3D DESS (respectively SAG/CORIW TSE images). An overview of the feature extraction process is presented in Fig. 2.

Following-up preprocessing, each slice of the input volume is passed to a pretrained model which will compute 1280 numerical descriptors (features), resulting in depth x 1280 descriptors (where depth corresponds to the number of input slices).

2.4.2 Architectures for 2D MRI sequences

Throughout this manuscript, 2D MRI sequences refer to both COR IW TSE and SAG IW TSE images listed in the OAI database. Those images usually contain less slices than pure 3D sequences such as SAG 3D DESS. Taking this into account, the architecture of the deep learning model used with 2D MRI sequences slightly differs from the ones developed for 3D sequences, as detailed thereafter.

Attention sub-model

In light of previous studies related to multiple instance learning [26], we first implemented a simple architecture whose aim was to compute attention scores (which can be viewed as “importance” scores) for each slice of the input image. Such scores were further used in the second part of the model. Starting from a set of one-dimensional feature vectors for each slice, a 1-dimensional convolution was applied (hence leading to one 2-dimensional matrix per image), followed by a Gated Recurrent Unit (GRU) layer. Such an architecture reduces the 2D matrix to a 1D vector (with one scalar score per input slice). This score was then scaled in the [0,1] interval through the use of a softmax activation function, thus preventing this sub-model to give full importance to all slices.

Classification sub-model

Following calculation of importance scores, a mean weighted by those scores was computed from all feature vectors. Consequently, the model learned to select slices carrying information through the attention sub-model, a global mechanism called “attention mechanism”. Clinical variables were standardized before being concatenated to this vector, resulting in a 1290-long vector. A description of clinical variables can be found in the additional materials, Table 1. This multimodal 1-dimensional vector was then fed into a multi layer neural network, followed by a softmax activation, outputting final class probabilities. In this approach, slices carrying little information (*e.g.* out-of-knee slices) were given low attention scores, hence participating little (or not at all) to the final logits computed by the second sub-model. A global overview of the model, f*rom a group of feature vectors (*one per slice of the image*) to the final prediction (e.g.* prediction of progression as an example*)* is presented in Fig. 3.

The purple-shaded area is a first sub-model aiming to locate regions of interest within input images. The green-shaded area represents the classification sub-model, which aggregates both image and clinical information into progression (or pain score) probabilities.

2.5 Human Benchmark

To further qualify the performance of our predictive model in the identification of 12 month-OA progressors, we undertook a comparative study with two expert radiologists, one senior and the other more junior, on the same task. The senior radiologist has specialized in musculoskeletal imaging for more than 20 years whereas the junior radiologist has 2 years of experience. We first selected 300 knee MRI with both SAG 3D DESS, 2D COR IW TSE and baseline clinical variables (age, gender, BMI, height, weight and minimum JSW in the medial compartment). These 300 knee images were then used to create 150 pairs of knee MRI, each pair being composed of both a 12 month-OA progressor and a non-progressor. To account for noise measurement on the minimum medial JSW, the 150 knee MRI of 12

month-OA progressors were chosen such that 10 knees were from “almost certain” 12 month-OA progressors with JSN(12 months) < -1.1mm, 130 satisfied − 1.1mm ≤ JSN(12 months) < -0.6mm and 10 were “doubtful” progressors with − 0.6mm ≤ JSN(12 months) < -0.5mm. These three classes of progressors reflect the distribution of 12 month JSN in the OAI population. In addition, the − 1.1mm threshold for “almost certain” 12 month-OA progressors was computed using the methodology from Parsons et al [27]. This threshold takes into account the standard deviation of JSW at baseline and 12 months and further ensures that, with a high probability (≥ 95%), the observed loss of knee cartilage is associated with a degenerative process rather than simply reflecting noise measurement. This methodology mimics the way ROC AUC is computed for a binary classifier [28] as it evaluates the ability of either the radiologists or the classifier to correctly rank two images (picked at random) knowing that one is a positive sample whereas the other one is a negative.

3. Results

3.1 Prediction of progression at 12 months

The model aims to identify knees for which JSN(12 months) ≤ -0.5mm. The five curves correspond to the five-fold cross-validation scheme. The dotted diagonal line (purple) illustrates the performance of a random predictor.

To identify 12 month-OA progressors from *baseline* knee MRI, we developed models to predict JSN(12 months) ≤ − 0.5mm from 2D (SAG IW TSE and COR IW TSE) as well as 3D (SAG DESS) knee MRI sequences. The most promising results were obtained using the classification model depicted in Fig. 3, which takes as an input 8 consecutive slices (centered around the “middle” slice) from a 2D COR IW TSE volume as well as the patient Body Mass Index (BMI) and predicts whether the observed knee is a 12 month-OA progressor, *i.e.* JSN(12 months) ≤ -0.5 mm. We thus report below on classification results obtained with 2D COR IW TSE sequences.

The performance of our classification model was evaluated using the ROC AUC score, well suited to this task as a metric given the class imbalance: only 14% of the patients with COR IW TSE images at baseline (3236 patients ; 5709 COR IW TSE knee images) are 12 month-OA progressors. Using COR IW TSE images, the proposed classification model achieved a ROC AUC score of 65%. This model achieved a precision of 13% and a recall of 84%. The above results are further summarized in a confusion matrix, reported in Fig. 4.

3.2 Human benchmark

The two radiologists concluded that, for most pairs, their decision was virtually random. Both radiologists found that 2D COR IW TSE volumes were less useful than SAG 3D DESS. The junior radiologist obtained a ROC AUC score of 57.82% whereas the senior radiologist obtained 59.72%. This benchmark with human radiologists highlights the difficulty of identifying 12 month-OA progressors using only knee MRI and

clinical data at baseline. Nonetheless, these results show the added value of AI in assisting radiologists in a complex image analysis task.

3.3 Prediction of pain severity

We subsequently applied our machine learning approach to the prediction of pain contemporary to image acquisitions. The grading of pain quantified by the WOMAC score was organized into two sets of values, including WOMAC pain score ≥ 2 and WOMAC pain score < 1. The rationale behind this stratification is two-fold. On one hand, it reflects some clinical relevance in that pain scores below 2 are often identified as “no pain”. Furthermore, it facilitates a data-driven approach where independent models can be trained and evaluated using only clinical data from different ranges of values. Such models were found to perform better when considering two sets of values with the above mentioned orders of magnitude.

Using this approach, our predictive model for pain achieves a mean PR AUC of 66.8% (+/- 1%), a mean ROC AUC of 72.4% (+/- 1%) and a mean weighted-F1 score of 65.2% (+/- 1%). Corresponding ROC and PR curves obtained for each of the five training folds are shown in Fig. 5. For comparison, a random predictor would achieve a mean ROC AUC of 50% and a mean weighted-F1 score of 60%. Globally, the model demonstrated good capabilities to identify high-pain knees (*i.e.* produce a relatively low number of False Negatives), with however a tendency to misclassify non-painful knees (*i.e.* produce False Positives), as can be seen in the confusion matrix represented Fig. 5, left panel. The associated PR curve is given in the additional materials, Fig. 8.

Graphic representations are shown for the binary task of classifying sets of pain scores across the cross-validation process.

3.4 Model interpretability

In order to get a better understanding of model predictions, the GradCam [29] method was further used to visually pinpoint relevant characteristics within input images, as shown in Fig. 6. Yellow-colored regions were identified within the joint as the ones contributing with a high probability to the positive class, *i.e.* progression in the case of JSN progression prediction (Fig. 6, top row), and high WOMAC pain score in the case of pain prediction (Fig. 6, bottom row), respectively. Purple-colored regions did not contribute to high probabilities in the predictions. Interestingly, this analysis emphasized different regions of interest depending on the task. Specifically, JSN progression-related regions are highlighted by the model in the internal femoro-tibial compartment. In contrast, for pain prediction, areas of interest are rather located in the intra articular space, where effusion is observed in the case of congestive osteoarthritis.

The bottom row corresponds to prediction of JSN progression and the upper row to pain prediction. Yellow areas are the ones considered of high interest by the model: the more intense the yellow, the higher its contribution to a high score for JSN progression prediction (bottom row, coronal view) or severe pain classification (top row, sagittal view). All images are obtained from patient 9932578 (right knee).

4. Discussion

Predicting disease progression in knee OA is critical to identify patients more likely to benefit from DMOADs and further, to help selecting patients and defining treatment duration in clinical studies evaluating drug candidates [30]. In the present study, we thus developed a weakly supervised deep learning method to build up predictive models for OA progression at 12 months from baseline MR images. Further analyses were also conducted to predict pain grade evaluated by WOMAC from MR images and clinical data at the same visit.

Using COR IW TSE images, our proposed classification model achieved a ROC AUC score of 63%. The latter was comparable to the performance of trained radiologists, obtaining a ROC AUC score of 59.72%. To our knowledge, this is the first application of a weak supervised learning method to the prediction of knee osteoarthritis progression from baseline MRI. Although not shown, no improvement on performance was observed on prediction of progression when considering a 24 month follow-up. We also successfully designed a task to identify imaging features associated with pain, leading to a model achieving a ROC AUC score of 72%. This encouraging result is likely explained by the presence of synovial effusion in painful knees, very contrasted in images and thus easy to detect for a radiologist. Our results are consistent with a previous study relying upon siamese neural networks to analyse pairs of knees and predict pain with a high AUC (85.3%) [15]. This study confirmed that 86% of correctly predicted painful patients exhibited an effusion-synovitis within areas most associated with pain.

Deep learning methods are often described as “black-boxes”, referring to the lack of interpretability of their predictions. Interpretability can however be introduced in the form of “heatmaps” generated using a GradCam method [29] to highlight the relevant regions in the knee MRI used by the predictive model. In our study, such attention modeling of OA progression confirmed the importance of internal joint space, consistent with the fact that the Joint Space Narrowing is evaluated in this anatomic compartiment. The pain prediction model rather showed heatmaps focused on the intra-articular space, where cartilage, meniscal lesions, and effusion synovitis are observed. In future developments, other interpretability methods based on Generative adversarial networks (GANs) could be applied to generate synthetic imaging features reflecting pathophysiologic processes of interest in OA. Whereas GANs modeling the natural history of OA progression observed on knee radiographs have been developed [31] such studies remain to be done on MRI.

Other developments in AI-based image analyses could be considered to improve the predictive models obtained in our feasibility study. For example, whereas we used MRI as inputs for predicting an endpoint determined from knee X ray imaging, further studies could rather use MRI criteria as endpoints of progression in clinical trials of knee osteoarthritis. The latter could circumvent part of the complexity in identifying the future progression of OA based on the absolute joint space narrowing > 0.5 mm as an endpoint, *i.e.* a criteria difficult to quantify reproducibly, even more so in light of interpersonal variability in joint space measurement. In this regard, we investigated in a post hoc analysis, the use of a different criteria to characterize OA progression. Whereas our initial analyses have been based on JSN(12 months) $\leq$ -0.5mm, reflecting that a knee is a “12 months OA progressor” when the minimum JSW is reduced by 0.5mm in the medial compartment of the knee, we reasoned that this “absolute” criteria may not be

suited to knees with advanced OA at baseline. We thus considered as an alternative a “relative” criteria defined by: JSN(12 months) ≤ -25% relative to JSW(baseline), choosing the latter threshold as it ensures that the dataset has approximately the same class imbalance as with the “absolute” criteria. In this approach, a classification model based on 2D COR IW TSE and validated using a 5-fold cross-validation strategy obtained an average ROC AUC score of 80%, suggesting an interest in considering relative over absolute JSN reduction as an alternative endpoint of OA progression.

5. Conclusions

Our study further demonstrates the added value of deep learning [32] in musculoskeletal imaging. On 2D radiographs, previous studies have been successfully conducted for bone fracture detection [33], as well as automatic Kellgren and Lawrence Grading for knee OA [34]. Other studies on knee MRI showed strong performance on cartilage segmentation [35], as well as detection or grading of meniscal or anterior cruciate lesions [36]. All these studies relied upon “strong” labeling methods, requiring time-consuming manual image annotations by expert radiologists experts. In comparison, the deep learning approach developed herein is based on “weak” labels for machine learning tasks, *i.e.* relying on information not explicitly shown in images as targets for predictions. The latter further demonstrates the added value of deep learning in clinical practice as it applies to OA, with the promise of a convergence of intelligences between machines and radiologists in the interpretation of radiological images [37, 38]. The future in the field is likely one of a new era of augmented radiology.

Abbreviations

DMOAD: Disease Modifying Osteoarthritis Drugs

GRU: Gated Recurrent Unit

JSN: Joint Space Narrowing

JSW: Joint Space Width

KL grade: Kellgren & Lawrence grade

OA: Osteoarthritis

OAI: the Osteoarthritis Initiative database

WOMAC: Western Ontario and McMaster Universities Arthritis Index

ROC AUC: area under the Receiver Operating Characteristic

PR / PR-AUC: Precision Recall / area under the Precision-Recall curve

GAN: Generative Adversarial Networks

MR(I): Magnetic Resonance (Imaging)

ML: Machine Learning

Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

The dataset used in the current study are publicly available from the OAI database (The Osteoarthritis Initiative 2002 [cited 2021 Jan 26]. Available from: https://nda.nih.gov/oai/)

7.4 Competing interests

J.-B. S., R. D., P. H., D. C., J. D., T. C. and G. W. are employees at OWKIN. F. K.-G., A. L., M. P., R. G., C. G. and P. M. are employees as Servier. The authors declare no competing interests in relationship with this manuscript.

7.5 Funding

This study was funded by Servier.

7.6 Authors' contributions

The study was designed by D. C., T. C., G. W., C.G. and P. M. The IT setup was performed by J. D. J.-B. S., R. D. and P. H. were involved in data preparation. P.H., F.K.-G., M.P., R.G. provided medical expertise in the interpretation of model outputs. J.-B. S., R. D.,P. H. C.G. and P.M. wrote the manuscript, which was critically reviewed by all authors.

7.7 Acknowledgements

The authors thank Dorothée Piva for excellent assistance in preparing the manuscript.

References

1. Woolf AD. The bone and joint decade. strategies to reduce the burden of disease: the Bone and Joint Monitor Project. J Rheumatol Suppl. 2003;67:6–9.
2. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. The number of persons with symptomatic knee osteoarthritis in the United States: Impact of race/ethnicity, age, sex, and obesity. Arthritis Care Res (Hoboken). 2016;68:1743–50.
3. Vina ER, Kwoh CK. Epidemiology of osteoarthritis: literature update. Curr Opin Rheumatol. 2018;30:160–7.
4. Hermans J, Koopmanschap MA, Bierma-Zeinstra SMA, van Linge JH, Verhaar JAN, Reijman M, et al. Productivity costs and medical costs among working patients with knee osteoarthritis. Arthritis Care Res (Hoboken). 2012;64:853–61.
5. Dell'Isola A, Allan R, Smith SL, Marreiros SSP, Steultjens M. Identification of clinical phenotypes in knee osteoarthritis: a systematic review of the literature. BMC Musculoskelet Disord. 2016;17:425.
6. Karsdal MA, Christiansen C, Ladel C, Henriksen K, Kraus VB, Bay-Jensen AC. Osteoarthritis–a case for personalized health care? Osteoarthritis Cartilage. 2014;22:7–16.
7. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. Bone. 2012;51:278–88.
8. Ravaud P, Giraudeau B, Auleley GR, Chastang C, Poiraudeau S, Ayral X, et al. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol. 1996;23:1756–64.
9. Cooper C, Adachi JD, Bardin T, Berenbaum F, Flamion B, Jonsson H, et al. How to define responders in osteoarthritis. Curr Med Res Opin. 2013;29:719–29.
10. Mazzuca SA, Brandt KD, Schauwecker DS, Katz BP, Meyer JM, Lane KA, et al. Severity of joint pain and Kellgren-Lawrence grade at baseline are better predictors of joint space narrowing than bone scintigraphy in obese women with knee osteoarthritis. J Rheumatol. 2005;32:1540–6.
11. Halilaj E, Le Y, Hicks JL, Hastie TJ, Delp SL. Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative. Osteoarthritis Cartilage. 2018;26:1643–50.
12. Hunter DJ, Altman RD, Cicuttini F, Crema MD, Duryea J, Eckstein F, et al. OARSI Clinical Trials Recommendations: Knee imaging in clinical trials in osteoarthritis. Osteoarthritis Cartilage. 2015;23:698–715.
13. Foreman SC, Neumann J, Joseph GB, Nevitt MC, McCulloch CE, Lane NE, et al. Longitudinal MRI structural findings observed in accelerated knee osteoarthritis: data from the Osteoarthritis Initiative. Skeletal Radiol. 2019;48:1949–59.
14. Madan-Sharma R, Kloppenburg M, Kornaat PR, Botha-Scheepers SA, Le Graverand M-PH, Bloem JL, et al. Do MRI features at baseline predict radiographic joint space narrowing in the medial compartment of the osteoarthritic knee 2 years later? Skeletal Radiol. 2008;37:805–11.
15. Chang GH, Felson DT, Qiu S, Guermazi A, Capellini TD, Kolachalama VB. Pairwise learning of MRI scans using a convolutional Siamese network for prediction of knee pain. bioRxiv. 2019;463497.

16. Zhang Y, Nevitt M, Niu J, Lewis C, Torner J, Guermazi A, et al. Fluctuation of knee pain and changes in bone marrow lesions, effusions, and synovitis on magnetic resonance imaging. Arthritis Rheum. 2011;63:691–9.
17. OAI. The Osteoarthritis Initiative [Internet]. The Osteoarthritis Initiative. 2002 [cited 2021 Jan 26]. Available from: https://nda.nih.gov/oai/
18. Peterfy CG, Schneider E, Nevitt M. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. Osteoarthritis Cartilage. 2008;16:1433–41.
19. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16:494–502.
20. Bruyere O, Richy F, Reginster J-Y. Three year joint space narrowing predicts long term incidence of knee surgery in patients with osteoarthritis: an eight year prospective follow up study. Ann Rheum Dis. 2005;64:1727.
21. Reginster J-Y, Reiter-Niesert S, Bruyère O, Berenbaum F, Brandi M-L, Branco J, et al. Recommendations for an update of the 2010 European regulatory guideline on clinical investigation of medicinal products used in the treatment of osteoarthritis and reflections about related clinically relevant outcomes: expert consensus statement. Osteoarthritis and Cartilage. 2015;23:2086–93.
22. Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J Neurosci Methods. 2016;264:47–56.
23. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Front Neuroinform [Internet]. Frontiers; 2011 [cited 2021 Jan 26];5. Available from: https://www.frontiersin.org/articles/10.3389/fninf.2011.00013/full
24. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29:1310–20.
25. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009. p. 248–55.
26. Ilse M, Tomczak JM, Welling M. Attention-based Deep Multiple Instance Learning. arXiv:180204712 [cs, stat] [Internet]. 2018 [cited 2021 May 11]; Available from: http://arxiv.org/abs/1802.04712
27. Parsons C, Judge A, Leyland K, Bruyère O, Petit Dop F, Chapurlat R, et al. Novel approach to estimate Osteoarthritis progression – use of the reliable change index in the evaluation of joint space loss. Arthritis Care Res (Hoboken). 2019;71:300–7.
28. Muschelli J. ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. J Classif. 2020;37:696–708.
29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis. 2020;128:336–59.
30. van Helvoort EM, van Spil WE, Jansen MP, Welsing PMJ, Kloppenburg M, Loef M, et al. Cohort profile: The Applied Public-Private Research enabling OsteoArthritis Clinical Headway (IMI-APPROACH)

study: a 2-year, European, cohort study to describe, validate and predict phenotypes of osteoarthritis using clinical, imaging and biochemical markers. BMJ Open. 2020;10:e035101.

31. Schutte K, Moindrot O, Hérent P, Schiratti J-B, Jégou S. Using StyleGAN for Visual Interpretability of Deep Learning Models on Medical Images. arXiv:210107563 [cs, eess] [Internet]. 2021 [cited 2021 Mar 24]; Available from: http://arxiv.org/abs/2101.07563
32. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. Nature Publishing Group; 2015;521:436–44.
33. Lindsey R, Daluiski A, Chopra S, Lachapelle A, Mozer M, Sicular S, et al. Deep neural network improves fracture detection by clinicians. Proc Natl Acad Sci U S A. 2018;115:11591–6.
34. Thomas KA, Kidziński Ł, Halilaj E, Fleming SL, Venkataraman GR, Oei EHG, et al. Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks. Radiol Artif Intell. 2020;2:e190065.
35. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection. Radiology. 2018;289:160–9.
36. Pedoia V, Norman B, Mehany SN, Bucknor MD, Link TM, Majumdar S. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. J Magn Reson Imaging. 2019;49:400–10.
37. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. Nature Publishing Group; 2019;25:44–56.
38. Lincoln CM, Chatterjee R, Willis MH. Augmented Radiology: Looking Over the Horizon. Radiology: Artificial Intelligence. Radiological Society of North America; 2019;1:e180039.

Figures

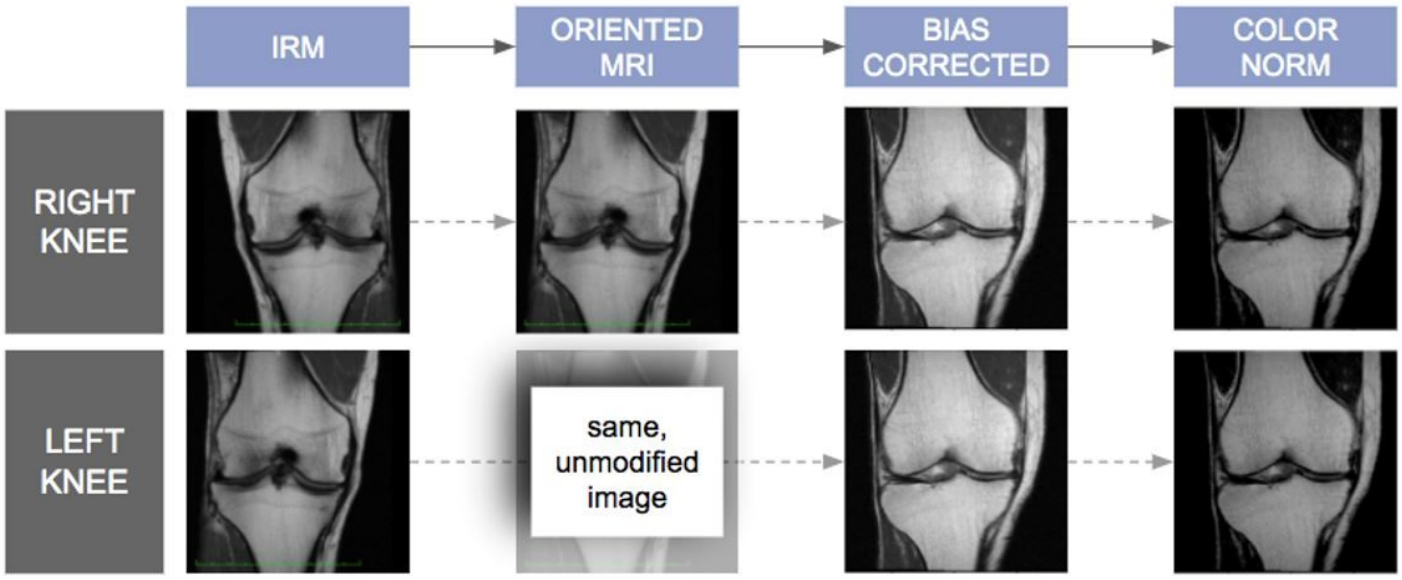

Figure 1

Overview of the image preprocessing pipeline.

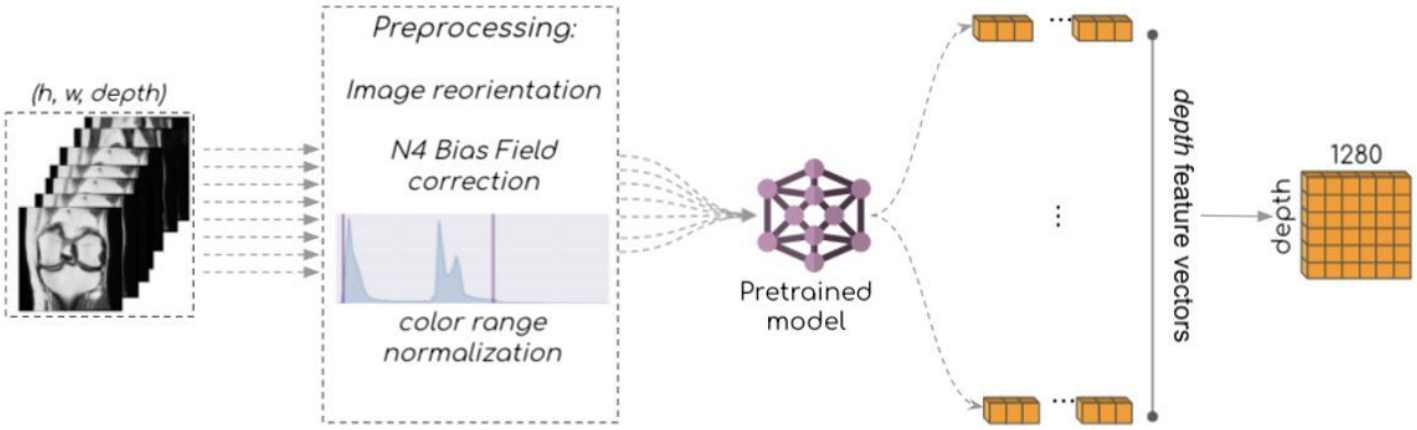

Figure 2

Global overview of the feature extraction step.

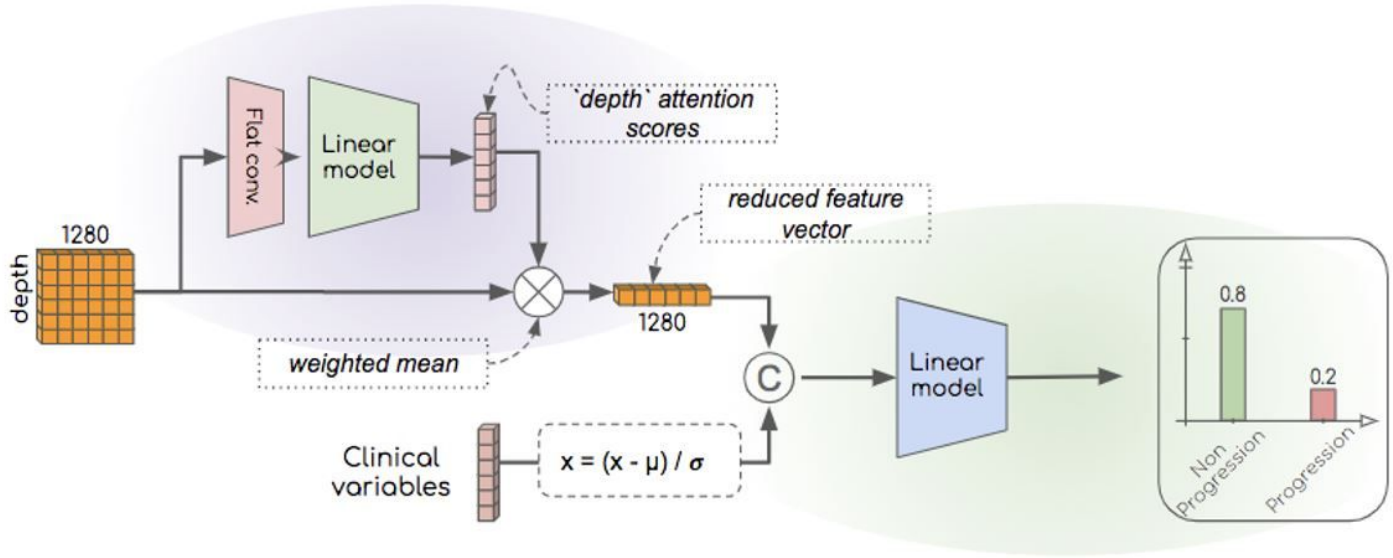

Figure 3

Global overview of the model.

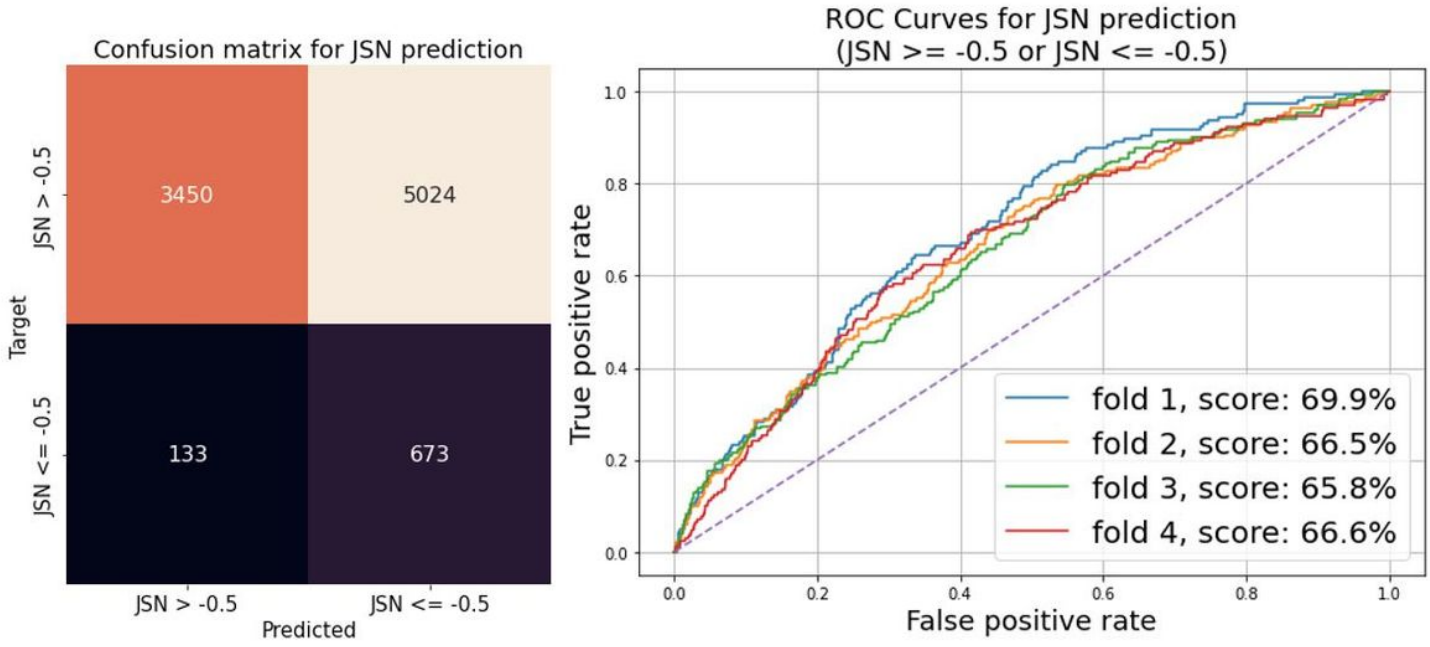

Figure 4

ROC curves and confusion matrix of the binary classification model to identify 12 month-OA progressors

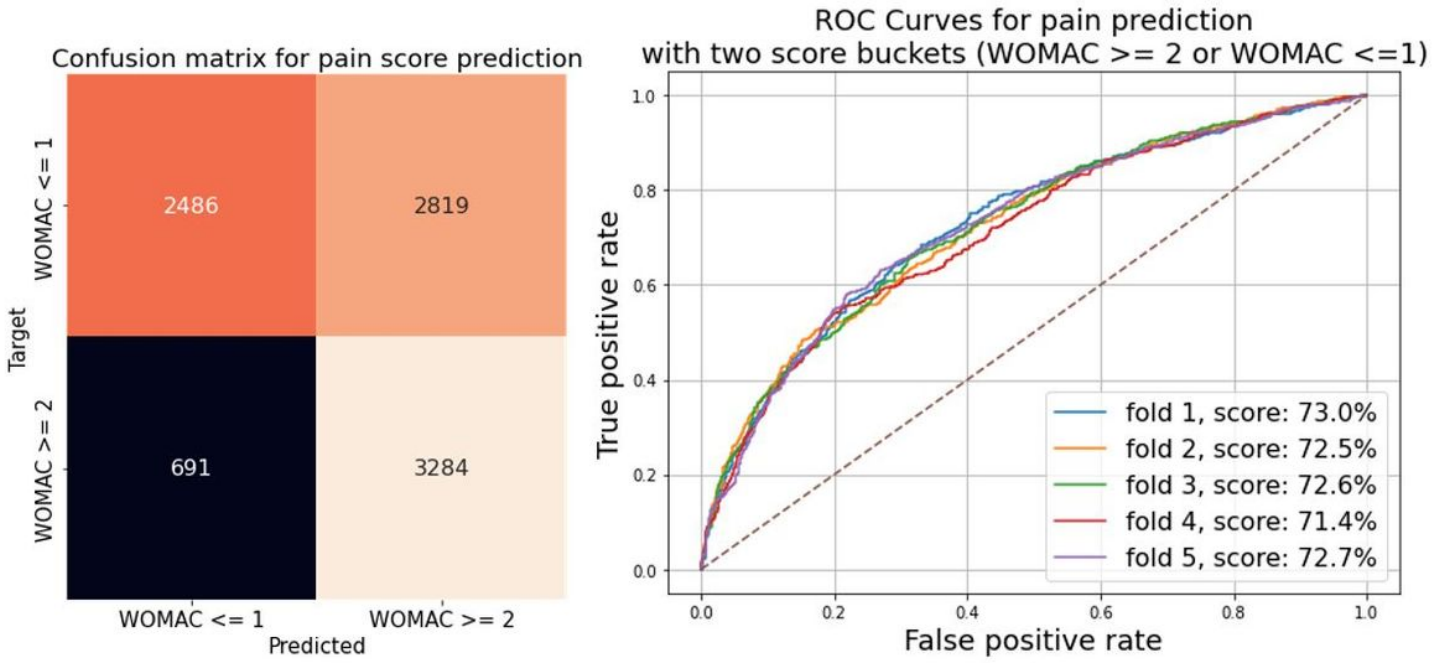

Figure 5

ROC curve and confusion matrix for prediction of pain severity. Graphic representations are shown for the binary task of classifying sets of pain scores across the cross-validation process.

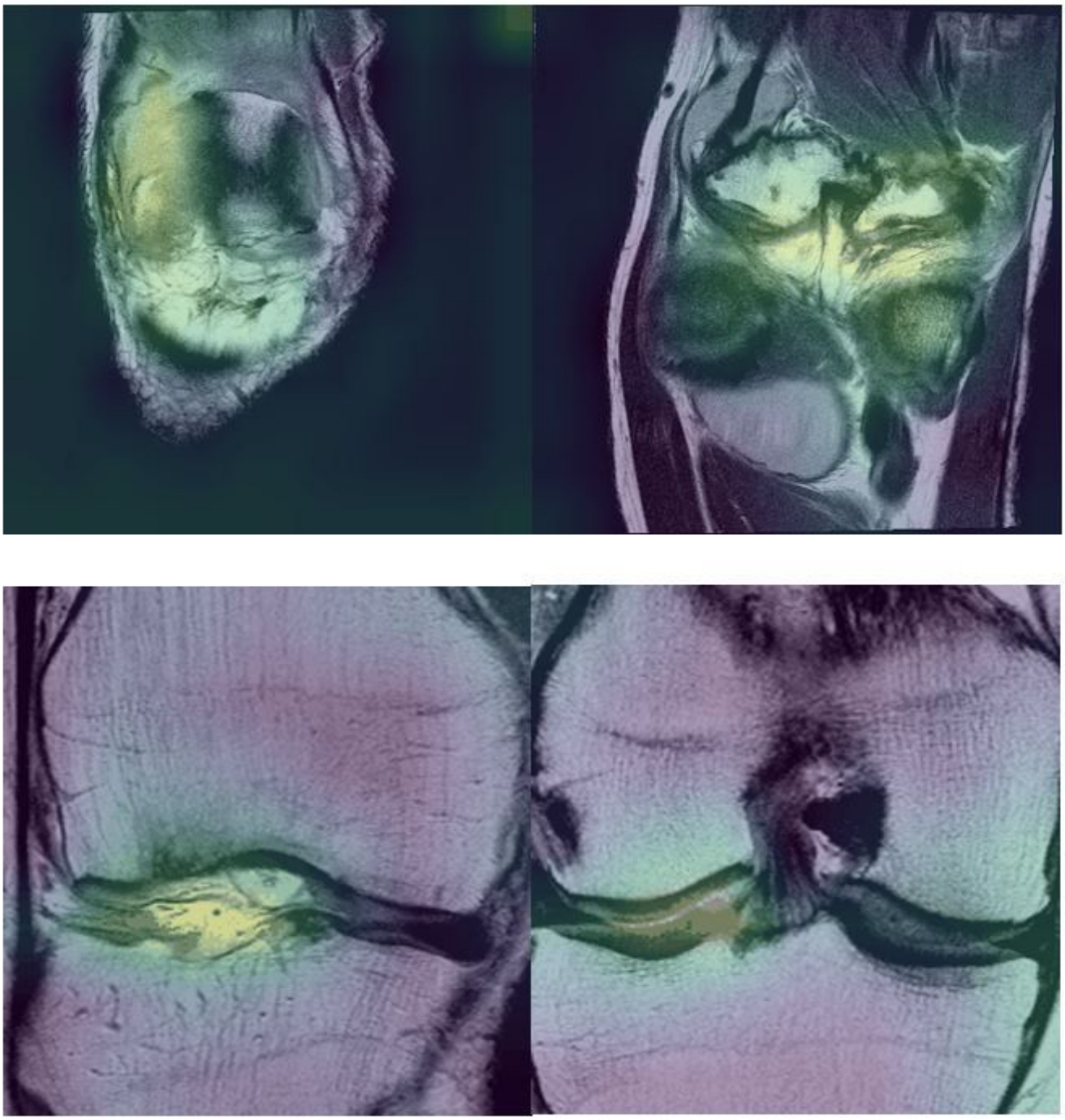

Figure 6

Visual interpretation of relevant zones identified by prediction models.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 03Additionalmaterials.docx