



Cite this: DOI: 10.1039/d5bm00470e

MRI detection and grading of knee osteoarthritis – a pilot study using an AI technique with a novel imaging-based scoring system

Chandrashish Roy,^a Mohammed Roshan,^a Nidhi Goyal,^b Prerana Rana,^c Nitin P. Ghonge,^b Amarnath Jena,^c Raju Vaishya^b and Sourabh Ghosh  [✉]

Precise and rapid identification of knee osteoarthritis (OA) is essential for efficient management and therapy planning. Conventional diagnostic techniques frequently depend on subjective interpretation, which have shortcomings, particularly during the first phases of the illness. In this study, magnetic resonance imaging (MRI) was used to create knee datasets as novel techniques for evaluating knee OA. This methodology utilizes artificial intelligence (AI) algorithms to identify and evaluate important indications of knee osteoarthritis, including osteophytes, eburnation, bone marrow lesions (BMLs), and cartilage thickness. We conducted training and evaluation on multiple deep learning models, including ResNet50, DenseNet121, VGG16 and ResNet101 utilizing annotated MRI data. By conducting thorough statistical analysis and validation, we have proven the efficacy of our models in precisely diagnosing and grading knee OA. This research presents a new grading method, verified by experienced radiologists, that uses eburnation as a significant indicator of the severity of knee OA. This study provides a new method for an AI-powered automated system designed to diagnose knee OA. This system will simplify the diagnostic process, minimize mistakes made by humans, and enhance the effectiveness of clinical treatment. Through the integration of AI-ML (machine learning) technologies, our goal is to improve patient outcomes, optimize the utilization of healthcare resources, and enable personalized knee OA therapy.

Received 27th March 2025,
Accepted 14th August 2025
DOI: 10.1039/d5bm00470e
rsc.li/biomaterials-science

1. Introduction

Knee osteoarthritis (OA) is a complex musculoskeletal and inflammatory disease marked by cartilage degradation, osteophyte formation, and synovial inflammation.¹ The clinical burden is significant, with many affected by symptomatic knee OA.² Symptoms include pain, joint stiffness, and reduced ability to perform daily activities.^{3–5} Knee OA also leads to structural changes in bone, synovium, and muscles, which worsen its impact. Earlier research has established a strong association between OA and aging, affecting a significant number of individuals aged 55 and older.^{6,7} The degenerative nature of knee OA is mainly due to mechanical erosion of articular cartilage, which has limited regenerative capacity,

presenting challenges in sustained healing.^{8–10} Early diagnosis is challenging due to the lack of visible early-stage knee OA signs, often detected only after significant joint changes, leading to chronic pain and functional restrictions.^{11,12} While physical exams and radiography provide pathophysiological insights, they typically do not measure early cartilage degeneration effectively.^{13,14}

The method of manually grading knee OA has been widely acknowledged as a challenging task, which is susceptible to errors. It is affected by difficulties like prejudice, low agreement between different users, and limited ability to reproduce results. Although pathologists possess high levels of knowledge, the subjective aspect of manual grading sometimes inhibits the detection of early or moderate stages of knee OA.¹⁵ To address these problems, our study aims to utilize artificial intelligence to transform the process of diagnosing and classifying knee OA. In this study, AI refers to the overarching field that involves creating intelligent systems capable of performing tasks that would typically require human intervention. Specifically, ML, a subset of AI, is used in our research to develop models that can learn from the MRI scan data to predict knee OA severity. We utilized deep learning, a further subset of ML, to build and train CNNs that can automatically

^aRegenerative Engineering Laboratory, Department of Textile and Fibre Engineering, Indian Institute of Technology Delhi, New Delhi-110016, India.
E-mail: Sourabh.Ghosh@textile.iitd.ac.in

^bIndraprastha Apollo Hospitals Delhi, Delhi Mathura Road, Sarita Vihar, New Delhi, India

^cPET SUITE (Indraprastha Apollo Hospitals and House of Diagnostics), Department of Molecular Imaging and Nuclear Medicine, Indraprastha Apollo Hospitals, Sarita Vihar, Delhi-Mathura Road, New Delhi 110076, India

extract relevant features from MRI images and classify knee OA parameters. In our present work, we went through an in-depth review of the many approaches used in the identification of knee OA, providing insight into their individual advantages, drawbacks, and difficulties. The scope of our research includes an array of traditional and innovative methods used for diagnosing knee OA, with the goal of gaining a comprehensive knowledge of their effectiveness in real-world medical settings.¹⁶ Although radiography is largely recognized as the most reliable method for diagnosing knee OA, its ability to detect early-stage knee OA and assess the development of the disease is limited.¹⁷ While MRI and computerized tomography scans are dependable techniques for detecting advanced knee OA, challenges in identifying the condition at an early stage are still encountered because of their limited capacity to detect soft tissues.¹⁸ MRI, although superior in visualizing soft tissues compared to CT, can be hindered by the resolution limitations and the need for highly specialized interpretation. A recent review by Kijowski *et al.* highlighted the growing use of deep learning in OA imaging, showcasing its potential for automated cartilage lesion detection, segmentation, and grading using MRI and X-ray data (*e.g.*, cartilage segmentation or single-feature analysis).¹⁹ Additionally, both techniques can struggle with differentiating between early knee OA changes and normal variations in knee anatomy, leading to potential misdiagnosis or underestimation of disease severity.²⁰ Furthermore, these imaging modalities are resource-intensive, requiring significant time and cost, and their accessibility may be limited in certain regions or healthcare settings.^{1,9} MRI has benefits in detecting knee OA at the initial stage, but earlier approaches such as the Whole-Organ Magnetic Resonance Imaging Score (WORMS) do not provide the necessary accuracy for precise grading. So, our research aims to overcome these limitations by utilizing advanced ML algorithms to create an automated model based on the MOAKS criteria.²¹ The MOAKS criteria evaluate the severity of knee OA using MRI scans, from which various factors are assessed such as cartilage condition, osteophytes formation, bone lesions, cysts, meniscal abnormalities, and ligamentous abnormalities.²¹ The grading of each characteristic ranges from 0 to 3 or 4, depending on the severity of documented disease.²² MOAKS provides a comprehensive and detailed semi-quantitative scoring system designed to assess these pathological features with precision. Bone marrow lesions (BMLs) are graded based on their size and presence of cysts, ranging from no lesion (grade 0) to lesions occupying more than 66% of the subregional volume (grade 3). Articular cartilage damage is scored according to the area and depth of the lesion, from normal thickness (grade 0) to full-thickness defects covering more than 50% of the area (grade 4). Osteophytes are evaluated by size in specific anatomical locations, from none (grade 0) to large (grade 3). This detailed and structured evaluation system provided by MOAKS ensures a fine-grained understanding of knee OA and its impact on knee joint structures, facilitating accurate monitoring of disease progression and treatment efficacy. By integrating these comprehensive assessments into an automated model,

our research aims to enhance diagnostic accuracy and treatment planning for knee OA.

In this study, we also used the DICOM (Digital Imaging and Communications in Medicine) format for storing and processing the MRI scans. DICOM is the standard format for handling, storing, and transmitting medical imaging data. It ensures that images, along with metadata such as patient information and imaging parameters, are properly stored and can be used in conjunction with different medical imaging software. DICOM files are commonly used in clinical practice and research to ensure compatibility and ease of analysis across various platforms. We used DICOM MRI scans to obtain high-resolution images of the knee joint, which were then pre-processed and fed into our machine learning models for knee OA detection and grading.

The objective of our approach is to precisely detect and classify knee OA, evaluate its intensity, and predict grading levels with improved accuracy. The study is motivated by the need for knee OA detection at an initial stage and the dataset used includes MRI scans from patients across all severity levels (grades 1 to 3). To demonstrate the model's diagnostic performance and clinical applicability, we have included a model-versus-clinical analysis for a grade 3 knee osteoarthritis condition in the Results section. Our strategy combines modern imaging techniques and AI-driven analysis to address the limitations of manual grading. This approach aims to achieve fairness, dependability, and enhanced diagnostic accuracy while also expanding our knowledge of the physiology of knee OA. The uniqueness of our research lies in the development of an innovative scoring system centred on eburnation, a key indicator of bone degeneration found in conditions like knee OA and non-union fractures. Traditionally, osteophytes have been regarded as the earliest radiographic indicator of osteoarthritis. However, through collaborative clinical validation with senior orthopaedic experts at Indraprastha Apollo Hospital, New Delhi, our study is the first to recognize eburnation as an even earlier and more definitive marker of knee OA onset. We observed that eburnation caused by direct bone-on-bone contact due to initial cartilage loss often precedes osteophyte formation, making it a crucial feature for early diagnosis. Its inclusion in our grading system allows for a more precise and progression-sensitive classification of knee OA. Eburnation reflects subchondral bone changes due to cartilage degradation. As cartilage wears away, the underlying bone undergoes sclerosis, forming a polished, ivory-like surface. This process, visible in MRI scans, serves as an early sign of knee OA before osteophytes appear. Eburnation is particularly useful in grading knee OA severity, as it directly correlates with bone adaptation to cartilage loss at an early stage. Previous studies have established its role in differentiating knee OA from other joint conditions supporting its inclusion in our grading system.²² Unlike existing approaches, our scoring system not only identifies the presence of eburnation but also quantifies its severity with high precision. This allows for detection of bone health issues at an initial stage and enables clinicians to predict patient outcomes more accurately.

Validated rigorously against established clinical standards, our system enhances diagnostic reliability and facilitates timely interventions, potentially improving treatment efficacy and patient quality of life.²³ This novel strategy uses the distinct clinical attributes associated with the advancement of knee OA, providing a targeted and accurate approach for early evaluation. Our focus is to use advanced imaging techniques and DL algorithms to address the few shortcomings of manual grading. This will result in faster and more accurate assessments, less bias in scoring, and improved dependability and repeatability. Furthermore, the final phase of our initiative was the development of DL-assisted automated grading models specifically built to assess human-knee MRI samples by measuring the extent and severity of disease progression. By substituting human grading systems, these models can eliminate differences between users, set a benchmark for knee OA rating, and provide more precise and efficient diagnosis. This research signifies an important advancement in the field of knee OA diagnosis, providing marked improvements in accuracy, efficacy, and unbiased judgment.

2. Methodology

We analysed MRI knee-scan data from a sample of 14 patients in collaboration with the Indraprastha Apollo Hospital, Delhi, India (ethical approval number-IAH-BMR-018/10-19). To establish a precise model for classifying knee OA (Table 1), each scan was meticulously examined. The assortment was carefully organised, highlighting significant indicators of the severity of knee OA. Specifically, a comprehensive analysis of the bone curvature in the MRI scans was required to detect eburation, which was a significant indicator of knee OA. The total of 14 patient samples in this study was divided into training, validation, and test datasets using an 80%–20% split, with 11 patients used for training and 3 patients reserved for testing. To support effective model training and ensure a balanced representation of disease severity, the dataset was randomly divided in a manner that guaranteed each set—training, validation, and testing—included cases from all three OA grades

(mild, moderate, and severe). Within the training set, we applied *k*-fold cross-validation (with *k* = 5) to ensure strong validation of the model. This means that the training set was split into 5 smaller subsets, and the model was trained and validated 5 times, each time using a different fold for validation and the remaining 4 folds for training. This process was repeated for each subset, ensuring that all data points were used for both training and validation. The final test set of 3 patients remained untouched throughout the training and validation phases and was only used for final model evaluation (Table 2).²⁴

After receiving the MRI scans, the dataset was organized and labelled based on the presence and the severity of knee OA. Each scan was examined to determine if knee OA was present and, if so, how severe it was. This classification and annotation process allowed for accurate labelling of the data, which was essential for training and evaluating the machine learning model.

The grading of the MRI scans using the MOAKS system was performed by an experienced and certified radiologist. The radiologist holds the required certifications in radiological interpretation and is regularly involved in clinical grading for knee OA at the hospital. The scoring process was supervised to ensure accuracy and consistency across all samples, and the grading system was validated by comparing the results with those of other experienced radiologists. The task involved categorising each image according to specific features indicative of knee OA, such as the presence of osteophytes, the formation of BMLs, eburation, and cartilage thickening.²⁵ The radiologists provided detailed scales for evaluating each parameter of the MRI scans. These scales were used to assess the presence and severity of knee OA in a standardized manner. By using these carefully defined evaluation criteria, the radiologists ensured that the labelling process was both accurate and uniform across all scans. This consistency was crucial for the subsequent training and evaluation of the machine learning model. Eburation, a characteristic feature of knee OA, involves significant changes in the subchondral bone, where increased osteoblastic activity leads to the deposition of sclerotic, ivory-like bone tissue.²⁶ As the cartilage thins and eventually exposes the subchondral bone, the bone undergoes adaptive changes such as increased mineralization and the formation of microfractures, resulting in a polished, hardened surface that not only increases joint friction but also contributes to chronic pain, inflammation, and reduced joint function through the formation of osteophytes and subchondral cysts.

Areas with the potential for eburation were identified by detecting deviations from the expected linear pattern. The pre-

Table 1 Overall quantity of acquired dataset sample images

Samples	Gender	Age (years)	Knee OA condition
Sample 1	Male	58	Mild
Sample 2	Female	47	Mild
Sample 3	Male	59	Mild
Sample 4	Female	69	Mild
Sample 5	Female	50	Mild
Sample 6	Female	45	Moderate
Sample 7	Female	46	Moderate
Sample 8	Male	54	Moderate
Sample 9	Female	54	Moderate
Sample 10	Female	56	Severe
Sample 11	Female	70	Severe
Sample 12	Female	47	Severe
Sample 13	Male	58	Severe
Sample 14	Male	46	Severe

Table 2 Total number dataset sample image collected

	Total images	Male	Female
Grade 1	511	329	182
Grade 2	1155	651	504
Grade 3	1099	630	469
Total	2765	1610	1155

liminary phase of our procedure involved improving the quality of the MRI scans in our dataset using essential preprocessing techniques.²⁷ These techniques included normalising the pixel values to establish a standard algorithm, resizing and cropping to ensure consistency, reducing noise to enhance clarity of the dataset, and adjusting the contrast to highlight important features according to requirements of the parameters. These procedures were designed to improve the clarity and consistency of bone structures, making it easier to analyse them with greater accuracy. Following that, we utilised edge detection methods to accurately outline the boundaries of bones in the dataset. By doing this crucial procedure, we were able to accurately identify and examine the contours of the bones with appropriate accuracy. This served as the foundation for finding the osteophytes and precisely determining the specific areas of eburnation, which are termed as a deviation from the bone's typical curvature. Following that, curvature analysis tools were used to determine the precise locations of eburnation in the scans. These tools helped identify and analyse deviations in bone curvature accurately.

The MRI scans were then examined to assess the presence of BML development by examining the colour grading and intensity of pixel colour blocks, which are indicative of differences in bone marrow intensity.²⁸ The regions corresponding to bone marrow were extracted using image segmentation methods.²⁹ The analysis involved quantifying the colour grading and intensity fluctuations using histogram analysis and intensity thresholding techniques. Regions indicating the development of BML were identified based on higher intensities.³⁰ The extent of cartilage thickness, which indicates knee osteoarthritic alterations, was assessed by comparing the thickness of the damaged and healthy sections. The MRI data were processed using image analysis techniques such as region segmentation and edge detection to isolate the cartilage layer.^{31,32} Following that, algorithms were developed to calculate cartilage thickness at various locations around the knee joint. This computational method involved measuring differences in thickness between healthy and impacted regions. To enhance model performance, regularisation and hyperparameter tuning were utilized during the optimization of cartilage thickness estimation.³³ To improve the model's accuracy, additional MRI knee scans were used. These scans were obtained from the same dataset of 14 patients, featuring different MRI sequences and views (such as axial, sagittal, and coronal) of the knee joints. These varied sequences enhanced the diversity and variability of the dataset, allowing the model to better segment and recognize key features such as osteophytes, BML formation and eburnations. The use of these additional scans improved the model's ability to generalize across different anatomical differences and disease manifestations, which ultimately contributed to more accurate predictions of knee OA severity.³⁴ Several ML techniques were used to create the knee OA classification model. The labelled dataset was utilised to train supervised learning algorithms such as Support Vector Machines (SVM), Random Forest, and Convolutional Neural Networks (CNN).^{35,36} In order to

enhance the performance of the models, regularisation and hyperparameter tuning were utilised during the optimisation process.³⁷ The performance of each model was evaluated using criteria such as F1-score, accuracy, and precision.^{38,39}

Deep Convolutional Neural Networks refer to neural networks with many layers (hence the term 'deep') that are capable of learning complex hierarchical features from input images. In our study, deep CNNs, such as ResNet50, ResNet101, DenseNet121 and VGG16, were used to automatically extract features from MRI images and classify knee osteoarthritis parameters. The Inception model is a type of CNN architecture that uses multiple convolutional filter sizes at each layer, allowing it to capture more varied features from the input images.⁴⁰ This approach helps the model learn from different scales of the data simultaneously, improving its performance in tasks like image classification. Custom-designed CNNs refer to specific architectures that were tailored to our dataset to enhance the model's ability to capture relevant features in the MRI scans. Unlike standard pre-trained CNNs, these custom models were adjusted and optimized based on our particular problem and dataset characteristics. Another technique, non-maximum suppression, is used during object detection to eliminate redundant bounding boxes and retain only the most accurate ones. This technique is crucial for detecting specific regions of interest, such as osteophytes, and ensures that the model does not produce overlapping or duplicate predictions.

2.1 Grading methodology

Our study utilized the MOAK system to evaluate the extent of knee OA by analysing the following factors: BMLs, osteophytes, eburnation, and cartilage thickness.²³ Each parameter was assigned a score ranging from 1 to 3 (Table 3).²¹ Table 3 presents a hybrid grading system that integrates MOAKS criteria (for osteophytes, BMLs, and cartilage thickness) with a new scoring framework for eburnation, which we developed based on clinical observations and prior literature. In our knee OA grading system, eburnation is emphasized as a key feature due to its pathological relevance in early-to-advanced disease progression. Its detection reflects cartilage degeneration that eventually contributes to subchondral remodelling and osteophyte formation, making it a reliable and underutilized marker of OA severity. Eburnation was quantified by analysing deviations in bone curvature and smooth surface formation, following methodologies similar to those described in previous imaging-based studies.⁴¹ The proposed scoring system for eburnation allows for a more standardized and objective assessment of subchondral bone changes, enhancing the accuracy of knee OA severity classification beyond traditional MOAKS-based evaluations.

Validation and reliability of our grading approach for eburnation in the scoring system were ensured through obtaining approval from a trained radiologist at Apollo Hospital. The scoring system was specifically developed to measure the gradual advancement of knee OA, where higher scores indicated a greater degree of disease development. This system

Table 3 Scoring categories based on different parameters for assessing the severity of knee OA

Parameters	Grade 1	Grade 2	Grade 3
Eburnations	2 blocks per 10 cm	2–7 blocks per 10 cm	>7 blocks per 10 cm
BML formation	10%	10–75%	>75%
Cartilage thickness loss	30%	33–65%	>65%
Binary classification			
	Yes	No	
Osteophytes	Presence of osteophytes	Osteophytes do not present in knee joints	

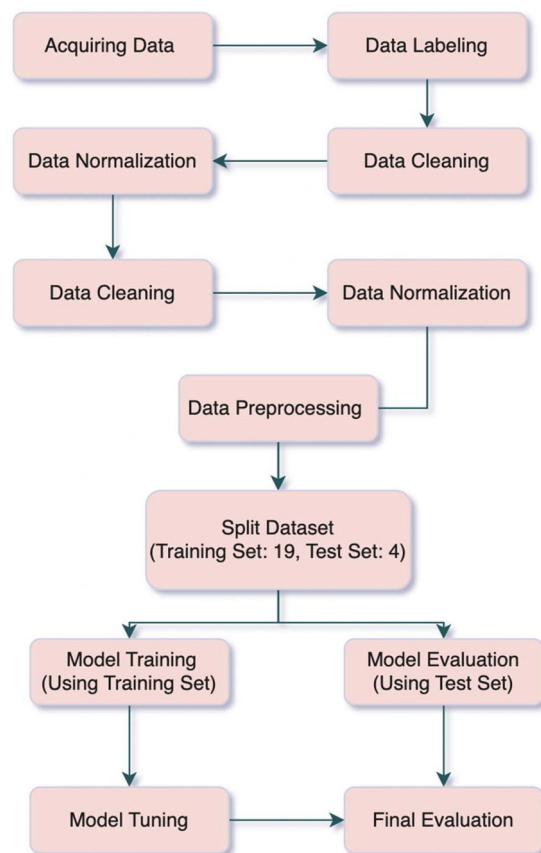
was designed to provide a detailed assessment of disease progression. A grade 1 score reflected the initial stages of knee osteoarthritic alterations in the examined parameter, indicating minor changes. In contrast, a grade 3 score suggested a significant level of disease, characterized by more evident alterations and more severe symptoms.

2.2 Model architecture

We utilized an advanced deep learning architecture specifically designed for image processing tasks to detect the targeted parameters present in MRI knee scans. The architecture comprised multiple components, each specifically designed to tackle distinct issues in the reliable and effective detection of osteophytes, eburnations, BML formations, cartilage thickness, and overall condition grading. We, then implemented an RPN within a Faster R-CNN framework to detect possible locations in MRI scans that contain osteophytes at the periphery of the knee regions and follow the algorithm to identify the required targeted parameters. The Reverse Polish Notation (RPN), which is smoothly integrated with the model backbone, effectively produces region ideas for subsequent analysis.⁴²

The approach consisted of a series of consecutive steps. First, data were collected from Apollo Hospital, and then subjected to thorough data preprocessing operations, which included data cleaning, labelling, and normalisation. Afterwards, the dataset was divided into separate test and validation sets. Subsequently, the model was trained using the training data, and its performance was thoroughly evaluated using the validation set. Modifications and fine-tuning of the model were conducted as needed, using the evaluation as a basis.⁴³ Finally, to make a definitive evaluation, the completed model was subjected to thorough examination using the separate test set of MRI scans for the 3 patients that was set aside and not used during the training phase. These scans were exclusively reserved for testing and provided an independent evaluation to assess the generalizability and accuracy of the machine learning models. The separate test set allowed for validation of the models on unseen data, ensuring that the models were not overfitting with respect to the training data and could reliably predict the classification of knee OA on new samples, ensuring the strength and applicability of the suggested methodology (Fig. 1).

The primary component of our model consisted of pre-trained CNN designs, such as VGG16, ResNet50 and

**Fig. 1** Project life cycle.

DenseNet121 (Fig. 2), which were incorporated into Region Proposal Networks.⁴⁴ This model detects and grades knee OA parameters from MRI scans. CNNs, including ResNet50, DenseNet121, and VGG16, were used for feature extraction and classification of knee OA indicators such as osteophytes, bone marrow lesions, eburnation, and cartilage thickness. These pre-trained CNNs, which were fine-tuned for our dataset, provided the backbone for the model's image analysis. Additionally, Region Proposal Networks (RPNs) were integrated with the CNNs to improve the detection of specific regions of interest, particularly osteophytes.⁴⁵ To further enhance classification performance, we also used traditional machine learning

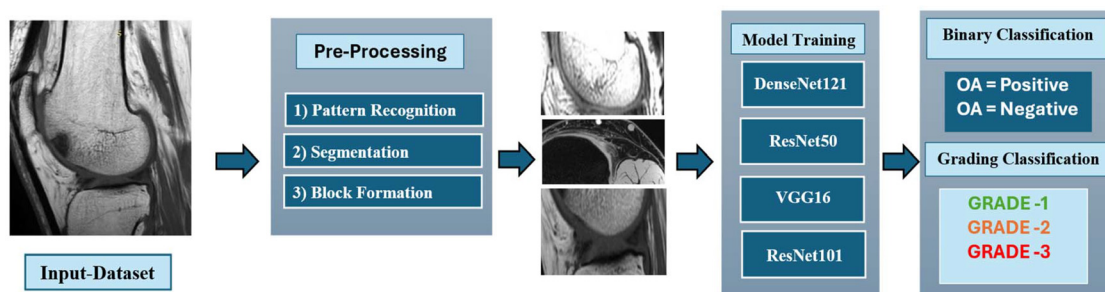


Fig. 2 Overview of the AI framework for knee OA detection. The model pipeline begins with CNN-based feature extraction using ResNet50, ResNet101, DenseNet121, and VGG16. These extracted features are further processed by Random Forest and SVM classifiers for precise knee OA grading. An RPN component is integrated to enhance the localization of knee OA-specific features in MRI images.

techniques, including Support Vector Machines (SVM) and Random Forest, for comparison and to verify the robustness of the results. These models worked together, with CNNs providing deep feature learning and SVM/Random Forest offering additional validation for the classification outcomes. This combination greatly improved our model's capacity to precisely detect and pinpoint the areas of interest in the dataset. Pre-trained models, which had been previously trained on large datasets like ImageNet, provided a strong foundation by extracting important features from scans. The extensive training of these models (Fig. 2) was used to recognize and understand various image patterns, making them highly effective for new tasks. By using these pre-trained models, we could apply their acquired information to our specific needs, reducing the amount of new data required and speeding up the training process. This approach ensured our models performed well and accurately on image-based tasks.⁴⁶

After localising the regions of interest, a deep CNN was used to perform comprehensive analysis and classification of the identified regions. We utilised architectures such as ResNet50, Inception, and custom-designed CNNs to detect our parameters, ensuring a balance between accuracy and processing efficiency.⁵¹ The utilization of the custom-designed CNNs, originally trained on large datasets, were leveraged for their powerful feature extraction capabilities. By fine-tuning these pre-trained models with our MRI knee dataset, our goal was to enhance the model's ability to identify specific signals critical for detecting and grading parameters such as BML formation, osteophytes, loss of articular cartilage lining, and eburnation (Fig. 3). This approach allows us to adapt the network to better recognize the unique features present in our MRI data related to knee conditions. To enhance the identification of the targeted parameters and guarantee precision, several post-processing procedures were employed. First, Non-Maximum Suppression (NMS) was used to eliminate redundant bounding boxes predicted by the model in the periphery knee region, ensuring that each osteophyte and instance of eburnation was recognised only once at the targeted sites.⁵² In addition, NMS facilitated accurate determination of BML presence and cartilage thickness based on the MOAKS scoring criteria.⁴⁷ Second, the use of Thresholding, which involved setting a confidence

threshold to exclude low-confidence detections, to improve the accuracy of osteophyte detection.⁴⁸

The model's performance was assessed using conventional metrics, including Precision, Recall, F1-score, and Intersection over Union (IoU), calculated specifically for bounding box predictions. An evaluation was performed on a distinct test set to appropriately assess the real-world performance of the model.

2.3 Patient selection criteria

The selection of patients for this study was based on well-defined inclusion and exclusion criteria to ensure the reliability and consistency of the dataset.

Inclusion criteria. Patients eligible for participation were those diagnosed with knee OA, confirmed through both clinical evaluation and radiographic imaging. This ensured that all selected cases met standardized diagnostic criteria, minimizing variability in disease severity assessment. To maintain high imaging quality and ensure accurate model training, only MRI scans with sufficient resolution and clarity were included, with a minimum requirement of 3T MRI scans. Additionally, patients had to be between 45 and 75 years of age, as this age range captures the most commonly affected population for knee OA while excluding younger individuals with post-traumatic knee OA and elderly patients with significant comorbidities that could interfere with imaging analysis.

Exclusion criteria. Patients with a history of knee surgery, trauma, or joint replacement were excluded to prevent structural alterations unrelated to knee OA from confounding the analysis. Additionally, individuals diagnosed with inflammatory arthritis, such as rheumatoid arthritis or other rheumatologic diseases, were not included, as these conditions could introduce pathological changes distinct from knee OA. Patients with other musculoskeletal disorders or significant joint deformities unrelated to knee OA were also excluded to maintain a homogeneous sample. Poor MRI image quality due to motion artifacts or technical limitations was another exclusion factor, as it could compromise image processing and model performance. Lastly, patients with incomplete clinical data were excluded to ensure a comprehensive dataset with complete demographic, clinical, and imaging records for accurate analysis and validation of the model.

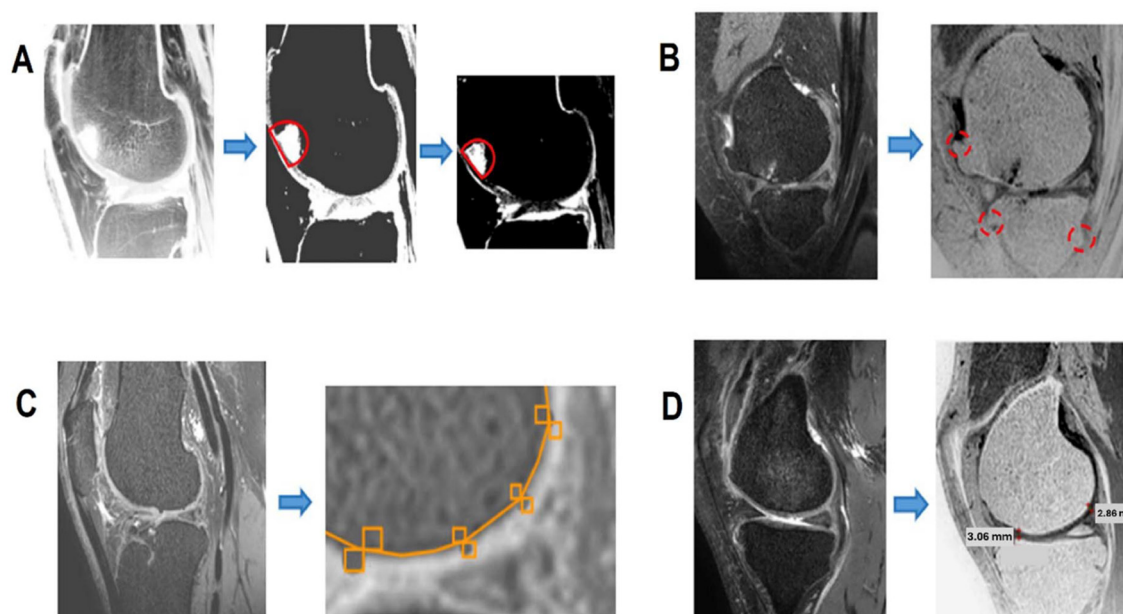


Fig. 3 Parameters for detecting the knee OA condition. (A) BML formation, (B) indicating osteophytes, (C) eburnation formation, and (D) loss of articular cartilage lining the medial tibial and femoral condyles.

2.4 Model training

The model training involved incorporating additional MRI knee scans to improve accuracy in detecting parameters by enhancing image segmentation and the pattern recognition method (Fig. 3). By initiating the different parameters of the dataset with a greater number of different knee MRI scans, the model was exposed to a wider variety of anatomical differences and pathological signs. This improved its capacity to accurately differentiate between normal and dysfunctional knee structures. Advanced image segmentation techniques along with pattern recognition algorithms were used to accurately outline specific areas of interest, such as osteophytes, within the periphery side of the knee in the MRI scans.⁴⁹ Utilizing the technique to build the algorithm to analyse the signals indicated the presence of osteophyte formations with increased sensitivity, as well as indicating the presence of BML and variations in cartilage thickness. The model-training process was conducted step by step, utilizing a larger dataset and advanced algorithms for segmenting and recognizing patterns.⁵⁰ This resulted in a more reliable and accurate tool for detecting osteophytes in MRI knee scans, with the potential to greatly improve the accuracy of diagnosing knee OA.

The eburnation detection block model played a crucial role in the ML architecture by identifying and analysing eburnation in MRI knee scans (Fig. 3C). This model utilized a combination of geometric analysis and computational techniques to track and assess the curvature of bone structures in the scan datasets. Specifically, it applied differential geometry-based curvature estimation, which enabled precise detection of irregularities along the bone boundary—such as bulges or indentations—that signified eburnation. To enhance accuracy, the model incorporated multi-scale feature extraction through a CNN,

which systematically analysed variations in bone morphology at different levels of detail. The curvature assessment was further refined using Gaussian curvature and shape descriptor algorithms, allowing the detection of subtle deviations from normal bone structure. This method ensured that the model could accurately delineate affected regions and quantify the extent of bone damage, providing valuable insights into knee OA progression. A standardized MRI-based grading method was tailored for knee OA detection at the initial stage. This system leveraged texture-based analysis and DL classification to assess cartilage degradation and subchondral bone changes with higher sensitivity than conventional scoring methods.⁵¹ The purpose of this system was to offer a more detailed evaluation of the severity of knee OA. Our methodology entailed tracking the curvature of the bone in the MRI scans to detect eburnations, which are a distinctive and the early-stage characteristic of knee OA.⁵² To measure the degree of eburnation, we utilized a technique based on blocks. Every recognized eburnation was delineated as a distinct unit, and the total count of these units within a 10 cm interval was recorded. This count was used to determine the severity of knee OA. A count of 2 blocks per 10 cm corresponded to grade 1, 2 to 7 blocks indicated grade 2, while more than 7 blocks indicated grade 3, as mentioned in Table 3. By utilizing this methodical approach, the diagnosis of knee OA was improved in terms of both accuracy and precision, which in turn allowed for more effective treatment planning and patient care.

2.5 Statistical analysis

The statistical analysis was conducted using R, a comprehensive statistical computing environment. The dataset consisted of MRI knee-scan records from a sample of 14 patients, which

were systematically divided into multiple sets to evaluate the performance of the deep learning models ($n = 3$).

This separation of data ensured that model performance could be assessed on unseen data, allowing for a reliable evaluation of generalizability. The study employed 5-fold cross-validation on the training dataset to optimize model learning and minimize overfitting. For statistical validation, $n = 3$ refers to the number of independent test samples used for the final evaluation of the trained models.⁵³ These three test cases were completely excluded from the training process to ensure an unbiased assessment. This approach helped in measuring the model's robustness, reliability, and ability to classify the severity of knee osteoarthritis accurately.⁵⁴

To quantify model accuracy, the following statistical metrics were applied:

- **Accuracy** = (true positives + true negatives)/total predictions
- **Precision** = true positives/(true positives + false positives)
- **Recall (sensitivity)** = true positives/(true positives + false negatives)
- **F1-score** = $2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$
- **ROC-AUC (Receiver Operating Characteristic-Area Under Curve)** to measure model discrimination ability.

To evaluate the accuracy of the classification models, particularly for categorising knee OA grades into four classes, various statistical markers were employed. These metrics included the calculation of average accuracy scores and F1 scores.³⁹ The accuracy scores provided a thorough assessment of the accuracy of the models in predicting knee OA grades (Table 4). However, the accuracy and recall metrics assessed the ability of the models to correctly classify positive cases and prevent false positives, respectively.

3. Results

3.1. Deep learning model comparison

During our comparative research of DL models, specifically VGG16, ResNet50, and DenseNet121, we thoroughly examined their ability to reliably identify knee OA parameters. By utilising these refined and enhanced CNN structures, we trained and assessed these models using a broad collection of MRI scans of knees that exhibited different levels of knee OA severity. Heat maps (Fig. 5) visually display the accuracy of each model in recognising specific knee OA metrics, such as osteophytes, eburnation, BML development, and cartilage thickness.⁵⁵ After conducting a thorough analysis using multiple models, it became clear that ResNet50 surpassed VGG16 and

DenseNet121 in terms of accuracy across all evaluated parameters. ResNet50 consistently achieved the highest accuracy rates (Fig. 5B). This is due to its effective use of residual connections, which addressed the vanishing gradient problem and allowed for deeper network training. To better understand the internal functioning of our deep learning system, Fig. 4 illustrates the detailed architecture of the ResNet50 model used in this study. Fig. 4 illustrates the complete workflow of our AI-based knee OA classification system, starting from data pre-processing to the final model output. On the left, the Data Pre-Processing stage shows how MRI slices are manually annotated using bounding boxes to isolate critical regions such as joint spaces, cartilage interfaces, and subchondral bone area regions typically affected in knee OA. This step is semi-manual and benefits from expert radiological input to ensure anatomical precision before feeding the images into the model. The image for the Model Integration section depicts how the cropped regions are processed by the ResNet50 architecture. Core CNN layers including convolution, activation, and pooling extract feature relevant to knee OA pathology, such as bone contour irregularity, cartilage thinning, and osteophyte formation. Supporting visuals highlight how the model detects and encodes structural patterns across layers. The final Output section displays the model's prediction performance. A probability curve illustrates class-wise prediction confidence, while a scatter plot presents the spatial separation of key knee OA features—osteophytes, eburnation, bone marrow lesions (BMLs), and cartilage thickness—as recognized by the trained model. DenseNet121 followed closely, benefiting from its dense connectivity that improved feature reuse and gradient flow, leading to high accuracy but still slightly behind ResNet50. VGG16, on the other hand, lagged behind both ResNet50 and DenseNet121. Its simpler architecture, while easier to implement, struggled with very deep networks, resulting in lower accuracy. As a result, ResNet50 proved to be the most reliable model for our dataset, with DenseNet121 performing well but not as effectively, and VGG16 showing respectable yet comparatively lower accuracy.

The accuracy of each model was calculated based on the proportion of correct predictions, defined as true positive/total predictions. In the corresponding figure, the X-axis represents different combinations of model parameters used during training, such as learning rate and batch size, while the Y-axis displays the true knee OA severity values. The values within the matrix indicate the accuracy of each model in predicting the actual severity of the disease across various knee OA features. The confusion matrices (Fig. 5B, C, and D) for ResNet50,

Table 4 Metrics of various models used for grading

Model	Grade 1 (%)	Grade 2 (%)	Grade 3 (%)	Average accuracy %	Precision	Recall	F1 score
ResNet50	89.775	80.828	89.453	86.685	0.74	0.72	0.7316
DenseNet121	83.987	79.987	86.345	83.440	0.72	0.71	0.7124
VGG16	78.876	71.546	78.451	76.291	0.69	0.67	0.6786
ResNet101	88.120	78.990	87.560	84.890	0.73	0.715	0.7267

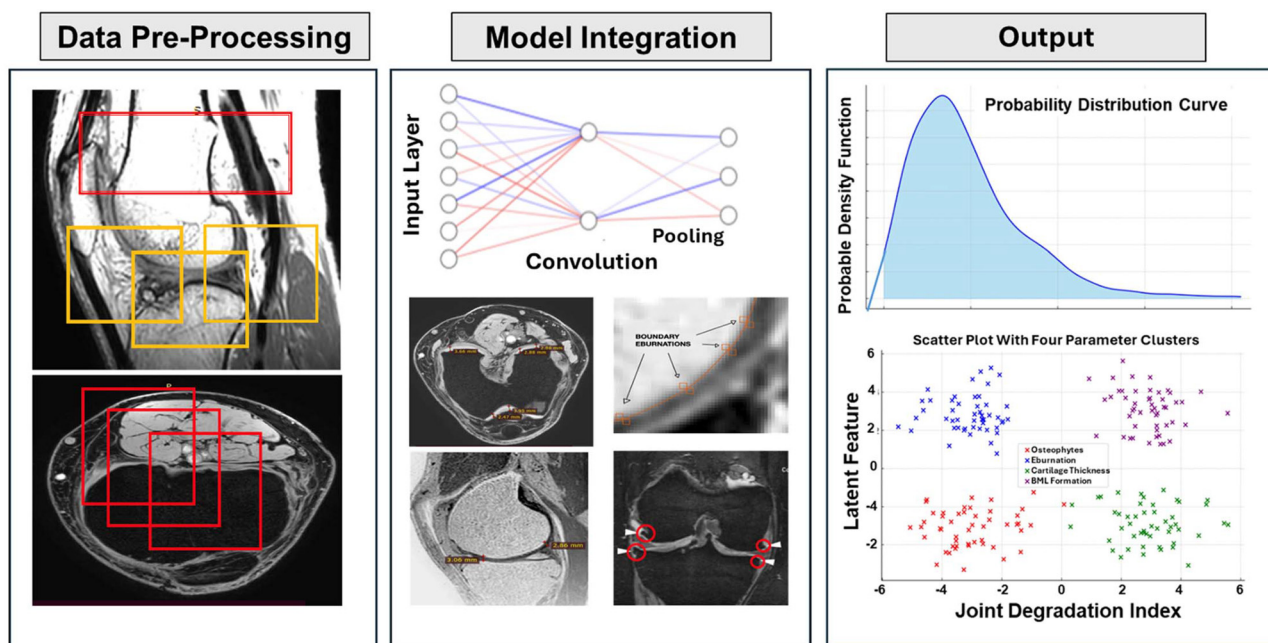


Fig. 4 Deep learning-based knee osteoarthritis assessment from MRI. The MRI images (left) are pre-processed by segmenting regions of interest. A convolutional neural network (middle) extracts features such as cartilage thinning and eburnation. The model outputs (right) include a probability distribution of joint degradation and clustering of knee osteoarthritis biomarkers. A detailed description of the ResNet50 architecture used for knee osteoarthritis detection and grading is provided in the text.

DenseNet121, and VGG16 provide a detailed performance comparison. These matrices showcase the ability of the models to predict key knee OA features, including BML formation, cartilage thickness, eburnation, and osteophytes. The diagonal cells represent correct classifications, whereas the off-diagonal cells indicate misclassifications. This visualization offers valuable insights into the classification performance of each model and its ability to accurately assess knee OA severity.

Aside from the heat map analysis, accuracy, epoch, and loss graphs offer further insights into the performance of each model during the training phase. Examining the accuracy graphs provides insight into how the categorization accuracy of each model changes across consecutive epochs. ResNet50 demonstrates a consistent improvement in accuracy, ultimately achieving the best accuracy rates compared to the other two models. Meanwhile, DenseNet121 exhibits a considerable increase in accuracy, but it is slightly lower than that of ResNet50. VGG16 exhibits inferior accuracy compared to both ResNet50 and DenseNet121 throughout the training epochs, albeit demonstrating some progress.

Furthermore, the epoch and loss graphs visually demonstrate the convergence and optimisation of each model throughout the training process.⁵⁶ The epoch graph of ResNet50 demonstrates a gradual and consistent convergence, suggesting effective learning and continuous improvement of the model over time (Fig. 6A). On the other hand, the epoch curve of DenseNet121 shows a slower rate of convergence, indicating that more time is needed for training to achieve optimisation (Fig. 6B). Meanwhile, VGG16, known for its consistently

slow progress over epochs, demonstrates poor learning when compared to ResNet50 and DenseNet121 (Fig. 6C). ResNet50 consistently shows a decrease in loss values, which suggests that the optimisation and convergence of the model are effective. DenseNet121 demonstrates a comparable pattern, but despite this, does so with marginally elevated loss values, indicating its slower rate of convergence. On the other hand, VGG16 exhibits varying loss levels, suggesting less consistent training dynamics and probable difficulties in achieving convergence.

The Accuracy and Loss graphs, as shown in Fig. 6, are presented for the three models used in this study: (A) ResNet50, (B) DenseNet121, and (C) VGG16. Each model's performance is shown over 50 epochs, with Accuracy on the left and Loss on the right. In Fig. 6A (ResNet50), the accuracy improves steadily, while the loss decreases rapidly after a few epochs. Fig. 6B (DenseNet121) shows a similar trend, though the accuracy increases with slight fluctuations, and the loss curve declines gradually. Fig. 6C (VGG16) demonstrates a similar pattern, with the accuracy reaching a peak before slightly stabilizing, and the loss decreases sharply. These graphs provide insights into how each model learns over time and the convergence of their training processes, with a noticeable decrease in loss for each model as the epochs progress.

3.2 Automated analysis of parameters using a deep learning model

Confusion matrices were generated to visually depict the classification results of the models. The matrices present the

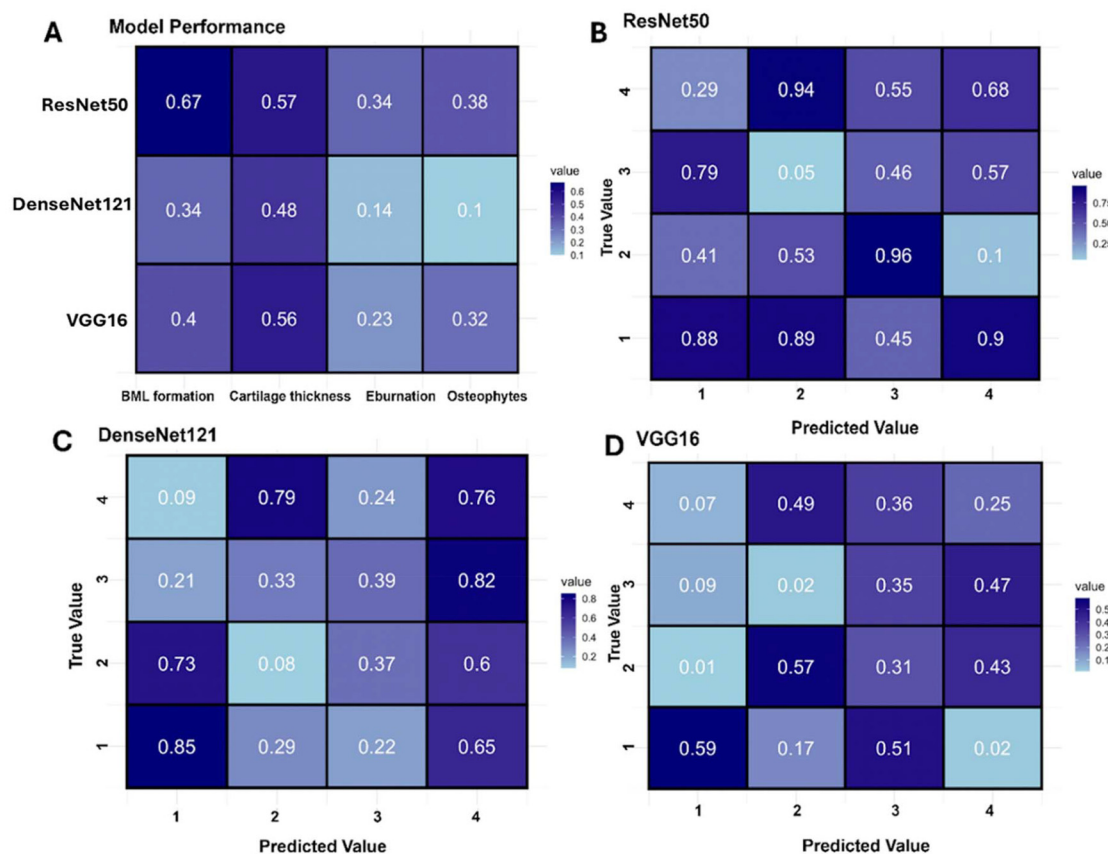


Fig. 5 (A) Matrix indicating that the model's accuracy fluctuated when different parameters were analysed, including factors like learning rate or batch size. (B) Graphical representation showing ResNet50 model accuracy across various parameter configurations, using colours to denote different accuracy levels. (C) Similar to (B), this heat map displays DenseNet121 model accuracy across parameter settings, aiding in identifying performance patterns. (D) A visual representation depicting VGG16 model accuracy under different parameter configurations, facilitating performance analysis.

number of true positives, true negatives, false positives, and false negatives for each class. ROC curves were created to evaluate the balance between the rate of correctly identified positive cases and the rate of incorrectly identified positive cases, and between the accuracy of positive predictions and the completeness of positive predictions, respectively. The areas under these curves (AUC) provided a quantitative measure of the each model's discriminatory abilities.

The statistical experiments were performed using the *R* environment, exploiting its powerful and available libraries and functions for ML and statistical modelling. The results of this research offer vital insights into the efficacy of the established models in accurately detecting and classifying knee OA. This information can be utilised in clinical settings to optimise disease diagnosis and improve patient treatment. Thus, scatter plots subsequently depict the prediction precision of the machine learning models, namely VGG16, DenseNet121, and ResNet50, within a technical framework. Each plot visually represents the performance of the model by showcasing the link between the true and expected responses. The metrics of osteophytes, eburnation, cartilage thickness, and BML for-

mation are essential indicators used in the diagnosis of musculoskeletal diseases (Fig. 7). Insights into the precision and consistency of the models can be obtained by analysing the alignment of data points along the regression line and the dispersion around it. Increased concentrations of points closely packed around the regression line indicate improved predictive accuracy, whereas larger dispersion indicates places where the model could be refined. Moreover, the differences in the percentage values allocated to each parameter indicate subtle variations in the performance of the model, which contribute to a thorough assessment of their usefulness in analysing medical scans. The best fit model with an accuracy of 86.68% was the ResNet50 model followed by DenseNet121 and VGG16 (Fig. 7D).

3.3 ROC curves

The Receiver Operating Characteristic (ROC) curves offer a comprehensive assessment of the models in detecting our targeted parameters (Fig. 8). These curves depict the relationship between the true positive rate and the false positive rate, providing a full assessment of the model's capacity to differentiate

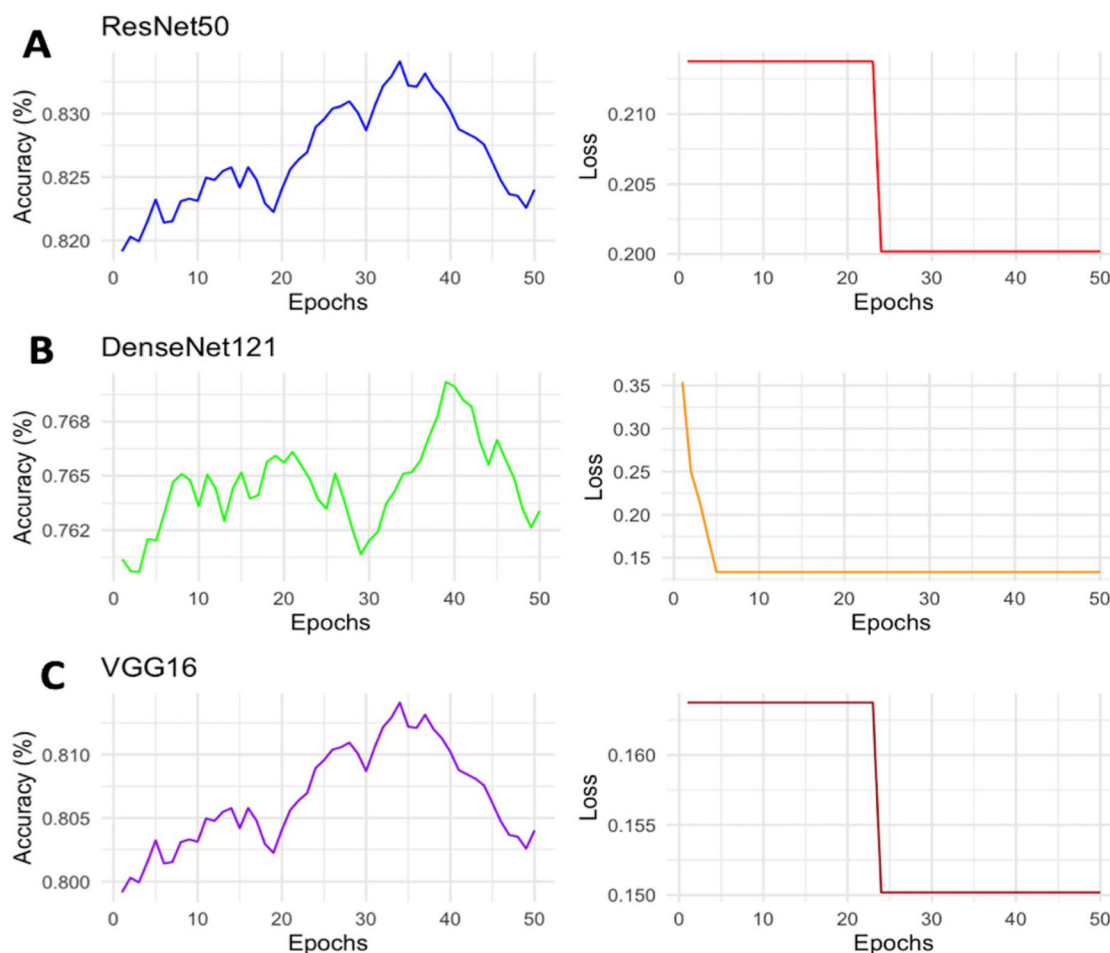


Fig. 6 (A) Accuracy and loss graphs for ResNet50, (B) accuracy and loss graphs for DenseNet121, and (C) accuracy and loss graphs for VGG16.

between various groups, which is crucial for accurate diagnoses. A model's capability to properly identify osteophytes improves as the area under the receiver operating characteristic curve (AUC-ROC) increases.⁵⁷ The sensitivity and specificity metrics highlight the effectiveness of these models by showing a well-balanced compromise, significantly reducing both incorrect positive and incorrect negative results. ROC curve evaluations are highly crucial to selecting the most efficient models for improving and maintaining an appropriate diagnosis accuracy. The accuracy metrics offer more quantitative understanding of the performance of each model in the task of detecting BML formations, cartilage thickness, eburnation and for detecting the number of osteophytes.

3.4 Does model architecture influence MRI-based knee OA grading? A comparison of ResNet101 and ResNet50

After testing with three standard models, we further compared the performance of ResNet101 with ResNet50 to assess whether increasing model depth could enhance the accuracy of knee OA grading, particularly in detecting critical pathological features. As illustrated in Fig. 9, this comparison includes classification heat maps (confusion matrices), scatter plots of

predicted *versus* actual feature values, and a bar graph comparing overall classification accuracy. While both models performed reasonably well across OA grades, ResNet50 demonstrated more prominent diagonal clustering in its confusion matrix (Fig. 9B), indicating more accurate and consistent classification than ResNet101 (Fig. 9A). The overall accuracy comparison shown in Fig. 9C reveals that ResNet50 achieved 86.68%, slightly outperforming ResNet101, which achieved 84.89%, indicating a performance improvement of +1.79%. Scatter plots (Fig. 9D and E) show predicted *versus* true values for four key OA-associated parameters: osteophyte formation, eburnation, cartilage thickness, and bone marrow lesion (BML) formation. While both models demonstrated strong correlation, ResNet50 predictions aligned more closely with the diagonal line, indicating higher precision. Notably, BML detection and cartilage thickness loss, which are often difficult to identify due to their subtle MRI features, showed the greatest performance gap between the models. ResNet101 achieved 76% accuracy for BML detection and 74% for cartilage thickness loss, whereas ResNet50 significantly outperformed it with 81% and 79% accuracy, respectively, highlighting its superior ability to detect these key OA features. In addition to compar-

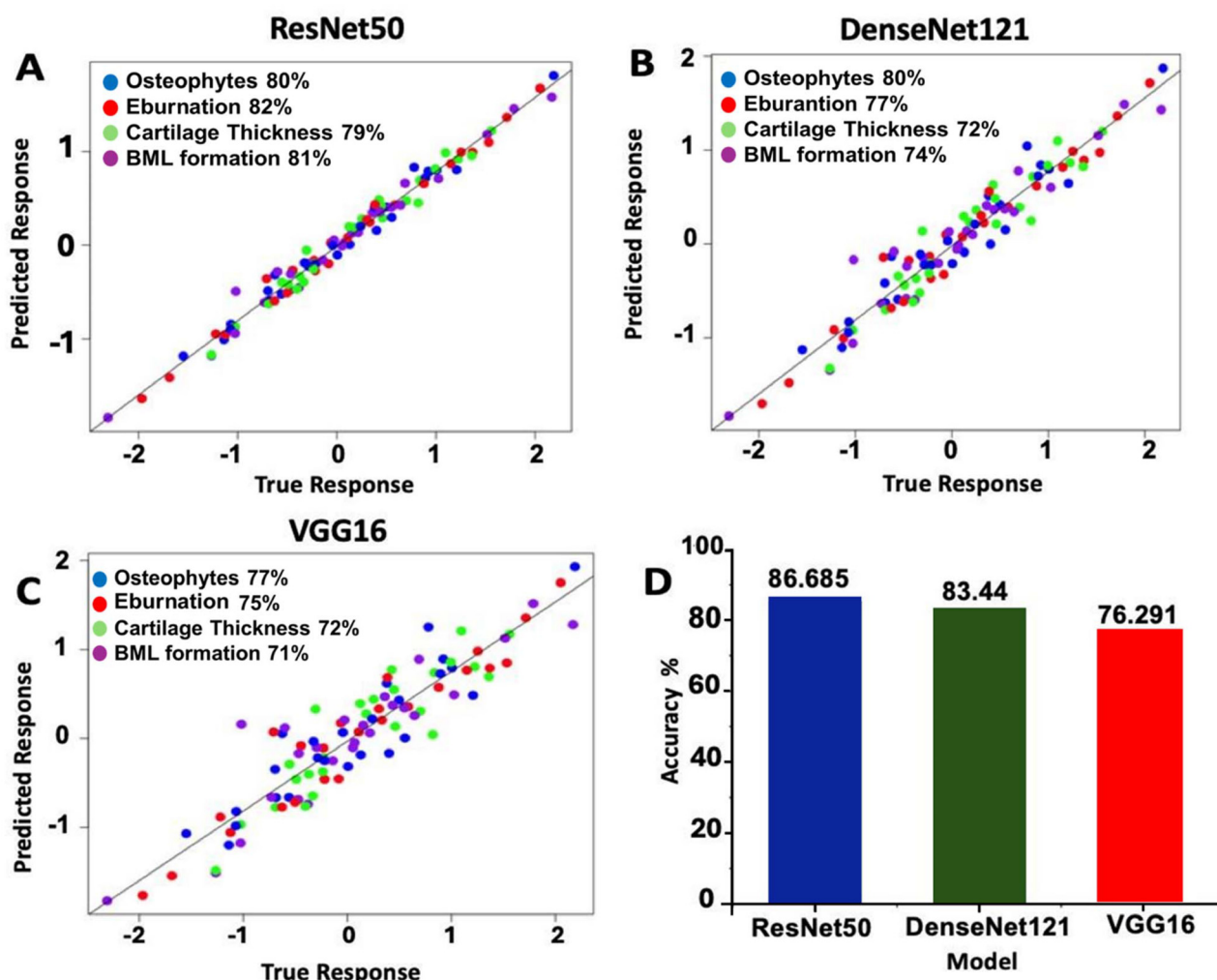


Fig. 7 (A) Scatter plot depicting predicted response versus true response for all four parameters using the ResNet50 model. (B) Scatter plot illustrating predicted response versus true response for all four parameters using the DenseNet121 model. (C) Scatter plot showcasing predicted response versus true response for all four parameters using the VGG16 model. (D) Bar graph comparing the accuracy obtained by each model, with ResNet50 achieving 86.68%, DenseNet121 achieving 83.44%, and VGG16 achieving 76.29%.

ing ResNet architectures, Fig. 9F presents the overall classification accuracy of all four deep learning models—DenseNet121, ResNet101, ResNet50, and VGG16. This broader comparison confirms that ResNet50 achieved the highest accuracy (86.68%), highlighting its robustness and reliability for automated knee OA grading.

3.5 Trial assessment results of deep learning model ResNet50 for MRI knee scans of two sample patients

In our study, we conducted an analysis of MRI knee scans of two patients using our machine learning model, ResNet50, to interpret MRI data with enhanced accuracy. The objective was to evaluate four key knee osteoarthritic parameters: eburnation, osteophyte presence, cartilage thickness, and bone marrow lesion (BML) formation. The methodology involved processing the patient MRI scans through our trained model, which subsequently performed a detailed assessment of the

aforementioned parameters. The findings for patient 1 are shown as visual outputs in Fig. 10(A, B and C), while the results for patient 2 are illustrated in Fig. 10(D, E and F). For patient 1, the identification of osteophytes resulted in a binary classification outcome confirming their presence, with a standard error of 0.5345. BML formation was categorized as grade 3, with a coefficient presence of 81.178% and a standard error of 0.5123. The thickening of cartilage, classified as grade 3, with a coefficient presence of 82.87%, showed a standard error of 0.5123. Eburnation was detected as grade 2, with a standard error of 0.5871. The overall diagnosis of knee OA was determined as grade 3, with a standard error of 0.5849. These findings are depicted in Fig. 10C. Similarly, for patient 2, the presence of osteophytes was identified with a binary classification outcome, yielding a standard error of 0.5879. BML formation was categorized as grade 3, with a coefficient presence of 77.178% and a standard error of 0.5098. The thickening of

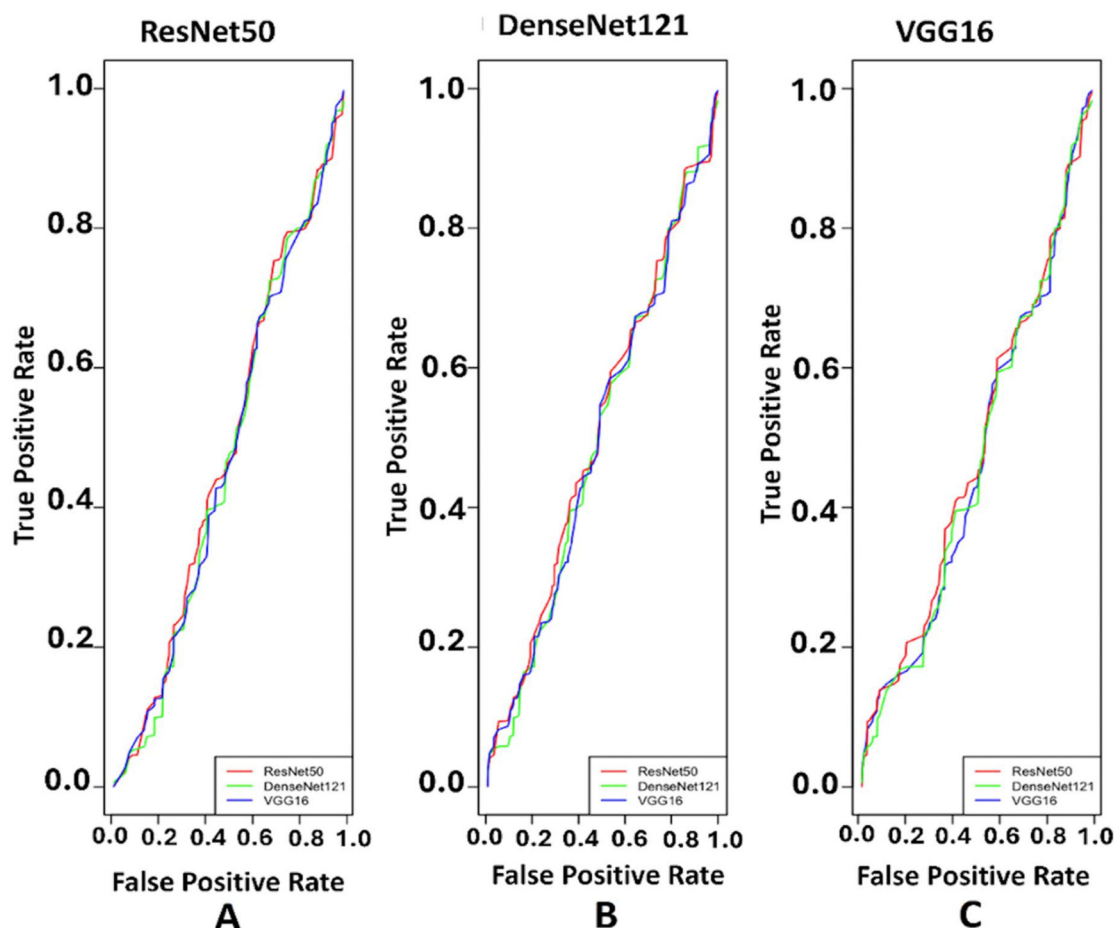


Fig. 8 (A) Accuracy curve: the accuracy of the ResNet50 model steadily increases with epoch values, reaching a peak accuracy of 99.15% at the end of training. Loss curve: the loss decreases consistently as the number of epochs increases, demonstrating the model's learning progress. (B) Accuracy curve: the DenseNet121 model shows a similar trend, with accuracy increasing over epochs, achieving a maximum accuracy of 97.16%. Loss curve: the loss decreases gradually as epochs progress, indicating effective learning by the model. (C) Accuracy curve: VGG16's accuracy also exhibits a positive correlation with epoch values, peaking at 81.78% after a certain number of epochs. Loss curve: the loss decreases steadily with increasing epochs, reflecting the model's ability to minimize errors during training.

cartilage, classified as grade 2, with a coefficient presence of 65.879%, exhibited a standard error of 0.5123. Eburnation was detected as grade 2, with a standard error of 0.5871. The diagnosis of knee OA for patient 2 was classified as grade 2, with a standard error of 0.6129. The model's output (Fig. 10F), aligns with these findings. These results confirm the model's ability to accurately interpret MRI scans and provide a strong framework for automated knee OA diagnosis.

3.6 Real-time comparison between the ResNet50 model and patient clinical data

We conducted a detailed analysis of MRI knee scans from a specific patient. The clinician performed an extensive manual analysis of the patient's scans, as shown in Fig. 11E, F, G and H. As mentioned in Table 7, the clinical analysis reveals that the patellofemoral joint has grade 3 knee osteoarthritis with significant BML formation, while the medial tibiofemoral joint has grade 2. The lateral tibiofemoral joint has grade 1 BML. In

the patellofemoral joint, cartilage thickness indicates a loss of approximately 60%, while it is 40% in the medial tibiofemoral joint, and 10% in the lateral tibiofemoral joint. It is observed that eburnation is 7 out of 10 cm. The results of our advanced AI model were precisely compared to the manual interpretation. This approach allowed us to evaluate the precision and reliability of our AI model against expert human assessment. The scans were carefully analysed by our model, which extracted important characteristics such as the presence or absence of osteophytes, BML, cartilage thickness, and eburnation. This analysis provided critical information on the patient's health. After assessment, our model accurately detected the existence and number of osteophytes, BML percentage, cartilage thickness percentage, and amount of eburnation with results of 4, 79.143%, 82.879%, and 8, respectively, as mentioned in Table 5 (Fig. 11A, B, C and D). The results of our study led to an overall grade of 3 for the patient (Fig. 11I), which is a severe condition, demonstrating the effectiveness of

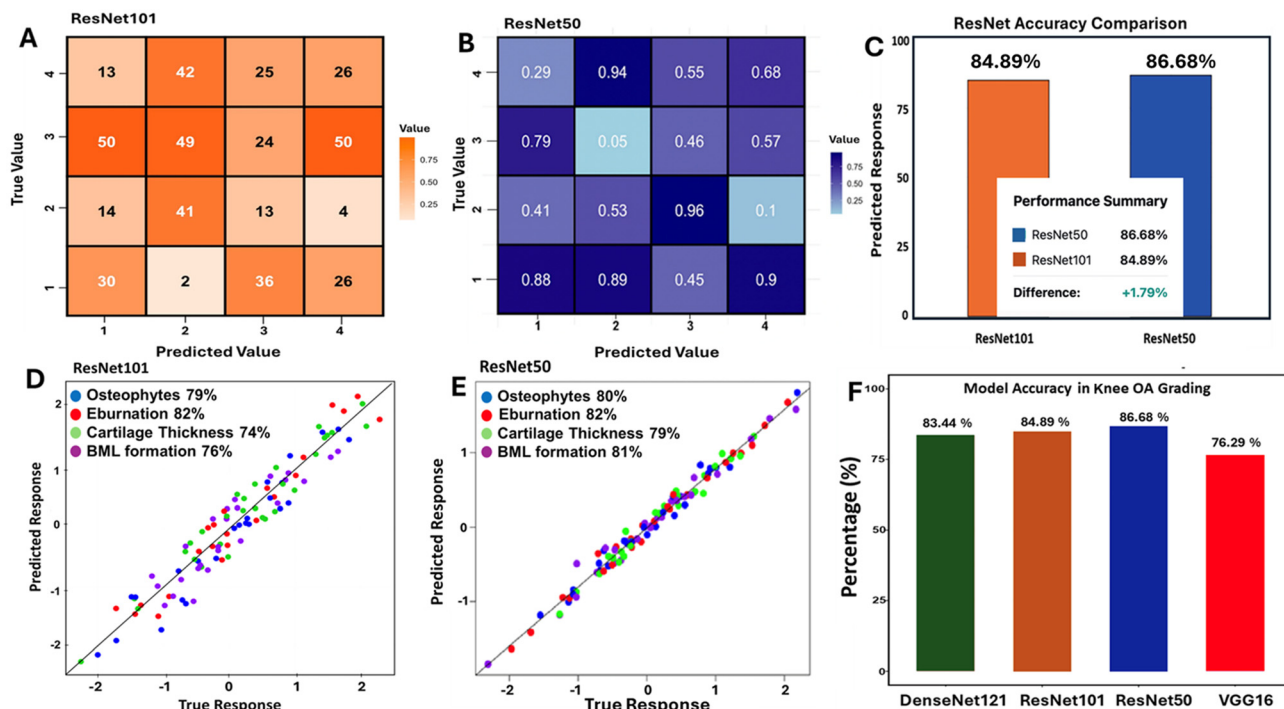


Fig. 9 (A) Heat map representing the confusion matrix of ResNet101 for OA grade classification. (B) Heat map showing the confusion matrix of ResNet50, indicating improved classification performance over ResNet101. (C) Bar graph comparing overall accuracy of ResNet101 (84.89%) and ResNet50 (86.68%). (D) Scatter plot of true vs. predicted responses for osteophytes, eburnation, cartilage thickness, and BML formation using ResNet101. (E) Scatter plot of true vs. predicted responses for osteophytes, eburnation, cartilage thickness, and BML formation using ResNet50. (F) Bar graph comparing model accuracy for knee OA grading: ResNet50 (86.68%), ResNet101 (84.89%), DenseNet121 (83.44%), and VGG16 (76.29%).

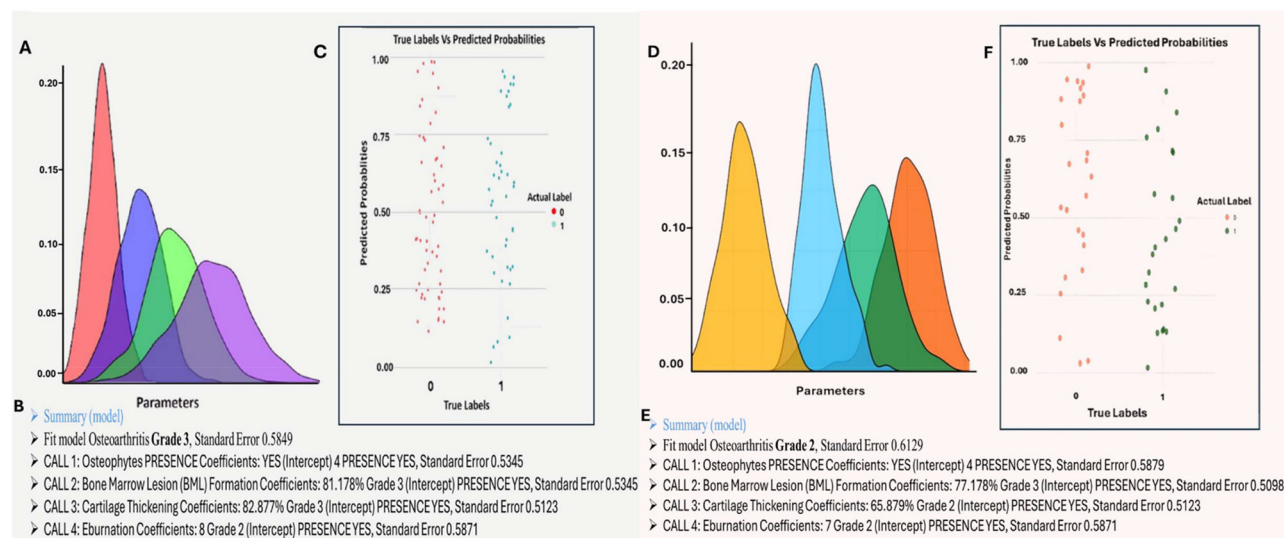
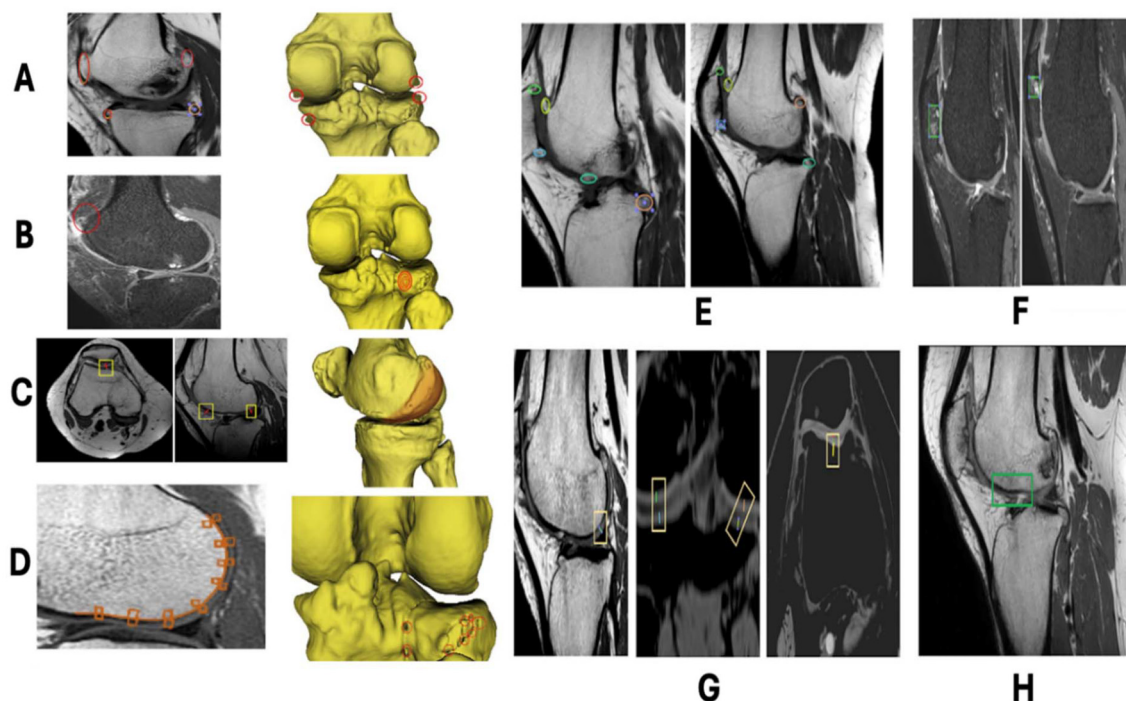


Fig. 10 Model-based analysis of knee osteoarthritis parameters for patient 1 (A, B and C) and patient 2 (D, E and F). (A and D) Probability density distribution of knee osteoarthritic parameters. (B and E) Model summary indicating knee OA classification grade and standard errors for osteophytes, BML formation, cartilage thickening, and eburnation. (C and F) Scatter plot of true labels vs. predicted probabilities, illustrating model classification accuracy.

our model in identifying important pathological characteristics with an accuracy rate of 84.326%. Furthermore, the manual evaluation conducted by the clinician produced identical grading results that closely matched the findings generated

by the ResNet50 model analysis, thus providing further evidence supporting the reliability and accuracy of our technique. To visually represent these findings, we utilized advanced software to develop the 3D knee of the DICOM MRI scans



Summary (model)

➤ fit model Osteoarthritis **Grade 3**, Std. Error 0.5849

➤ CALL 1 Osteophyte Presence Coefficients: YES (Intercept) 4 PRESENCE YES Std. Error 0.5345

➤ CALL 2 Bone Marrow Lesion (BML) Formation Coefficients: 79.143% Grade 3 (Intercept) PRESENCE YES Std. Error 0.498

➤ CALL 3 Cartilage Thickening Coefficients: 82.879 Grade 3 (Intercept) PRESENCE YES Std. Error 0.5671

➤ CALL 4 Eburation Coefficients: 8 Grade 2 (Intercept) PRESENCE YES Std. Error 8.5218

Fig. 11 Model analysis vs. clinical analysis of the parameters for detecting knee OA for a grade 3 (severe condition) patient. Model analysis of the parameters for detecting knee OA and reconstruction of 3D images from 2D MRI scans for a better understanding. (A) Osteophytes, (B) BML formation, (C) cartilage thickness and (D) eburation. Clinical analysis of the parameters for detecting knee OA for a grade 3 (severe condition) patient. (E) Osteophytes (different shape and colours representing the presence of osteophytes), (F) BML formation (green box), (G) cartilage thickness (yellow box) and (H) eburation (green box). (I) Overall model output results of the assessment.

Table 5 Clinical analysis: grade 1 knee OA

Parameters	Osteophytes	BML formation	Cartilage thickness loss	Eburation
Results	Present (6)	■ Grade 1 in patellofemoral joint; grade 1 in medial tibiofemoral joint ■ Grade 0 in lateral tibiofemoral joint ■ Grade 3 = approx. 20–30% of the entire bone area ■ Grade 2 = approx. 10–20% of the entire bone area ■ Grade 1 ≤ 10% of the entire bone area	■ Cartilage thickness loss in patella femoral joint is about 20–30% ■ Cartilage thickness loss in medial tibiofemoral joint is about 10–20% - Cartilage thickness loss in lateral tibiofemoral joint is nil	0–2 blocks in 10 cm diameter
Knee OA Grade			Grade 1	
ResNet50 model assessment				
Parameters	Osteophytes	BML	Cartilage thickness loss	Eburation
Results	Presence (4)	8.06%	27.733%	2 blocks in 10 cm diameter
Knee OA grade			Grade 1	

(Fig. 11A, B, C and D), effectively demonstrating the discovered analysis results. This visual depiction was obtained by using 3D Slicer software that is effective for clarifying the observed changes, providing more evidence supporting the reliable and accurate performance of our model. In addition to the grade 3 case, a comparative analysis was performed between clinical findings and the ResNet50 model for grade 1 and grade 2 knee OA (Table 5 and Table 6). For grade 2, the clinical assessment recorded osteophytes (6), bone marrow lesions (BML) involving 58.28% of the area with grade 2 changes in the patellofemoral and medial tibiofemoral joints and grade 1 changes in the lateral tibiofemoral joint. Cartilage thickness loss was approximately 30–60% in the patellofemoral joint, 30–40% in the medial tibiofemoral joint, and minimal in the lateral tibiofemoral joint, with an overall loss of 65.51%. Eburnation was observed in 3 blocks per 10 cm (average ~4 blocks per 10 cm). The ResNet50 model produced nearly identical findings, detecting osteophytes (5), BML of 58.28%, cartilage loss of 65.51%, and eburnation in 4 blocks per 10 cm, and classified the condition as grade 2. For grade 1, the clinical evaluation

showed osteophytes (6), BML of 8.06% with grade 1 changes in the patellofemoral and medial tibiofemoral joints and no involvement in the lateral tibiofemoral joint. Cartilage thickness loss was about 20–30% in the patellofemoral joint, 10–20% in the medial tibiofemoral joint, and no loss in the lateral tibiofemoral joint, totalling 27.73%, with eburnation limited to 0–2 blocks per 10 cm (average ~2 blocks per 10 cm). The ResNet50 model closely matched these results, detecting osteophytes (4), BML of 8.06%, cartilage loss of 27.73%, and eburnation of 2 blocks per 10 cm, and classified the case as grade 1, confirming the strong reliability between clinical and AI-based model assessments.

4. Discussion

Our research has presented an innovative method for automating the evaluation of human knee OA cartilage MRI scans utilising advanced DL techniques. This research aims to assess the viability of utilising DL for grading based on MRI scans,

Table 6 Clinical analysis: grade 2 knee OA

Parameters	Osteophytes	BML formation	Cartilage thickness loss	Eburnation
Results	Present (6)	■ Grade 2 in patellofemoral joint; grade 2 in medial tibiofemoral joint ■ Grade 1 in lateral tibiofemoral joint ■ Grade 3 = approx. 20–30% of the entire bone area ■ Grade 2 = approx. 10–20% of the entire bone area ■ Grade 1 ≤ 10% of the entire bone area	■ Cartilage thickness loss in patellofemoral joint is about 30–60% ■ Cartilage thickness loss in medial tibiofemoral joint is about 30–40% ■ Cartilage thickness loss in lateral tibiofemoral joint is minimal	3 blocks in 10 cm diameter
Knee OA grade			Grade 2	
ResNet50 model assessment				
Parameters	Osteophytes	BML	Cartilage thickness loss	Eburnation
Results	Presence (5)	58.28%	65.513%	4 blocks in 10 cm diameter
Knee OA grade			Grade 2	

Table 7 Clinical analysis: grade 3 knee OA

Parameters	Osteophytes	BML formation	Cartilage thickness loss	Eburnation
Results	Present (6)	■ Grade 3 in patellofemoral joint; grade 2 in medial tibiofemoral joint ■ Grade 1 in lateral tibiofemoral joint ■ Grade 3 = approx. 20–30% of the entire bone area ■ Grade 2 = approx. 10–20% of the entire bone area ■ Grade 1 ≤ 10% of the entire bone area	■ Cartilage thickness loss in patellofemoral joint is about 60% ■ Cartilage thickness loss in medial tibiofemoral joint is about 40% ■ Cartilage thickness loss in lateral tibiofemoral joint is about 10%	7 blocks in 10 cm diameter
Knee OA grade			Grade 3	
ResNet50 model assessment				
Parameters	Osteophytes	BML	Cartilage thickness loss	Eburnation
Results	Presence (4)	79.143%	82.879%	8 blocks in 10 cm diameter
Knee OA grade			Grade 3	

which represents a significant progression in the area, as opposed to previous research that mostly examined histology images and scan output from other methods, such as X-ray and CT analysis.⁵⁸ The study aims to use DL models to assess the severity and extent of knee OA by considering important indications of knee OA progression, such as the presence of osteophytes, automated grading of BML, thickening of cartilage, and the formation of eburnation. An important advantage of this technique is its capacity to mitigate bias and reduce disputes that can occasionally arise during manual grading procedures. The study resolved concerns regarding human subjectivity and variability in interpretation by implementing automated grading. This automation not only increases the effectiveness of grading but also promotes reliability and uniformity of the findings, which are crucial for precise diagnosis and prognosis of knee OA.

Expanding on the histology-based approach of Nagarajan *et al.*,⁵⁹ which utilizes the Modified Mankin scoring system for detecting knee OA, we propose an innovative methodology leveraging MRI-based deep learning techniques to automate knee OA grading. Our study seeks to address the limitations associated with the cost and accessibility of histological methods by developing an MRI-based model capable of accurately diagnosing and grading knee OA. This model promises a time-efficient and cost-effective alternative to manual diagnosis, incorporating factors such as osteophyte presence, cartilage thickness, and BML generation based on the MOAKS scoring system. Additionally, we are pioneering a scoring system for eburnation, an early sign of knee OA. Although histology offers precise knee OA assessment, it is prohibitively expensive, not universally accessible, and time-consuming. Therefore, our approach aims to create an accessible, cost-effective solution while maintaining the high accuracy and precision associated with histology. Unlike histological evaluation, which requires invasive tissue sampling and preparation, MRI is a non-invasive imaging technique that does not involve tissue extraction or cause any discomfort to the patient. This non-invasive nature makes MRI-based grading more feasible and less burdensome, allowing for precise monitoring of disease progression through repeated assessments. Although the primary focus of our study was on eburnation as an indicator of knee osteoarthritis, we also evaluated osteophyte identification due to its clinical relevance in knee OA diagnosis and progression. Osteophytes are widely recognized as a key feature of knee OA, and their detection provides valuable information for assessing disease severity.

Eburnation was chosen as the key classification feature based on expert validation from Indraprastha Apollo Hospital, which identified it as the earliest detectable marker of knee OA. Unlike osteophytes, which appear at a slightly later stage, eburnation reflects minute initial structural changes, making it a precise, innovative, and clinically important indicator in the AI-based knee OA grading system we have developed. So, by evaluating both eburnation and osteophytes, we aimed to provide a more comprehensive analysis of the model's capabilities in detecting multiple knee OA features. This dual

approach allowed us to assess the model's accuracy in identifying early signs of knee OA (through eburnation) and more advanced signs (such as detecting osteophytes), ensuring the robustness of the model across the spectrum of disease severity. Another study shows that while Kijowski *et al.* highlight the success of DL in OA lesion detection and segmentation, they also note the absence of models that integrate multiple joint features critical to disease progression. Their review noted that most DL models have been focused narrowly on cartilage or bone surface changes, and overlook early pathological markers such as subchondral remodelling. In contrast, our study incorporates eburnation—a key indicator of subchondral bone adaptation—alongside osteophytes, BMLs, and cartilage thinning within a radiologist-validated DL framework.¹⁹

Furthermore, in comparison with knee OA detection and severity classification using X-ray scans conducted by Mohammed *et al.*,⁶⁰ our MRI-based study offers several distinct advantages. First, MRI provides superior differentiation of soft tissues compared to X-rays, enabling improved visualization of structures such as cartilage, ligaments, and BMLs. This capability is crucial for diagnosing knee OA, as changes in cartilage thickness and BMLs are significant indicators of disease progression. Furthermore, MRI is free of radiation, enhancing patient safety, and its multi-planar imaging capability allows for a comprehensive evaluation of the knee joint, offering a detailed analysis of parameters not achievable with X-rays. By integrating artificial intelligence and machine learning technologies with MRI, our study aims to streamline the diagnostic process, minimize errors, and optimize treatment strategies, thus significantly advancing the field of knee OA diagnosis. This integration ensures that the diagnostic process is not only more accurate but also more efficient, ultimately benefiting both patients and healthcare providers. The findings of our study constitute a significant step forward in the field of artificial intelligence-driven healthcare, namely in the domain of diagnosing knee OA. Our model exhibits a high level of precision, with an accuracy rate of roughly 86%, even though there are still areas that need to be developed through further study. We utilized four different deep learning models for our project. Specifically, we tested ResNet50, DenseNet121, and VGG16. We also compared ResNet50 with a more complex model, ResNet101, to evaluate and check their overall performance (Fig. 9A, B, C, D and E). While both models gave more or less similar results, ResNet50 consistently performed better, especially in detecting bone marrow lesions (BMLs), with 81% accuracy compared to ResNet101's 76% as stated in the Results section. This performance gap highlights the limitations of using deeper models like ResNet101 when working with relatively small or limited datasets. Due to its greater number of layers and parameters, ResNet101 is more susceptible to overfitting—performing well on the training data but failing to generalize effectively to new, unseen data. In contrast, ResNet50, with its simpler architecture, was able to generalize better and produced more stable and reliable results across all evaluated features, making it a more suitable model

for our study. After thorough evaluation, ResNet50 demonstrated the greatest accuracy at 86%, outperforming ResNet101, which achieved 85%, DenseNet121, which achieved 83%, and VGG16, which achieved 76%. Following collaboration with radiologists from Indraprastha Apollo Hospital, Delhi, India and the incorporation of their expertise into the development of our model, we have prepared our approach to precisely meet the requirements of clinical practice. In order to identify OA by taking into consideration parameters such as eburnation, cartilage thickness, osteophyte presence, and BML formation, our model, which has been trained using MRI scans and fine-tuned to identify knee OA, has significant potential for altering the diagnosis of knee OA. This was shown in Fig. 4, which highlighted the clinical relevance and robustness of our deep learning pipeline using the ResNet50 model. The process starts with radiologist-provided MRI images, including both full views and zoomed-in sections that focus on regions commonly affected by knee osteoarthritis, such as cartilage interfaces, joint spaces, and bone surfaces. This targeted image selection ensures that the model is trained on high-quality, relevant anatomical features without requiring manual annotation or segmentation. Once fed into the system, the ResNet50 model effectively extracts and learns disease-specific patterns such as cartilage thinning, eburnation, and osteophyte development through its deep convolutional layers. We have incorporated the block formation, which gives an easy understanding when there is a formation of eburnation. Following that, we incorporated a method of Image Segmentation, which trained the model accurately for a better result survey of the MRI scans. Also, our algorithm followed a pattern recognition process based upon the MOAKS scoring system for accurately determining the grading. The more radiologists utilize our model, the more accurate and effective it becomes, benefitting more users in turn. When compared to manual grading systems, our automated DL classification approach has the potential to eliminate bias and inconsistency, both of which are present in such processes. Within the domain of knee OA, this system presents an opportunity to potentially enhance both research and clinical practice.

5. Conclusion

This study demonstrates a significant advancement in the diagnosis of knee OA, utilizing DL methods to automatically assess the severity of the condition using MRI scans. This research successfully addresses the subjectivity and variability that are present in traditional evaluation procedures by transitioning from manual to automated grading methods. The created deep learning model demonstrates robustness and dependability in assessing the severity of knee OA directly from MRI scans, eliminating the necessity for manual intervention. The model's performance and adaptability are improved by using a variety of datasets that include healthy and knee OA-affected cartilage from different MRI scans. The efficiency of the standard DL-based automated grading system

was validated using evaluation criteria such as accuracy curves, ROC curves, and scatter plots. This establishes the system as a standardized method for grading knee OA. This transformation offers the potential for enhanced efficiency, accuracy, and impartiality in the identification and diagnosis of knee OA. We found that incorporating deep learning-based automated grading systems into clinical practice has great potential to improve the management of knee OA. This technology revolutionizes patient care and treatment strategies by offering clinicians concise, unbiased, and reproducible estimates of disease severity using MRI imaging.

Ethical statement

All experiments were performed in accordance with the Guidelines of the Ethics Committee of Indraprastha Apollo Hospital, Delhi, India, and were approved by the ethics committee under approval number IAH-BMR-018/10-19. The patient sample data were obtained from Indraprastha Apollo Hospital with informed consent from all patients. Subsequent data processing and analysis were conducted at the Indian Institute of Technology Delhi, India.

Conflicts of interest

No competing financial interests exist.

Data availability

The clinical patient MRI dataset used in this study is not publicly available due to confidentiality and institutional restrictions. However, datasets may be provided upon reasonable request to the corresponding author, subject to ethical and regulatory approvals. The AI models and code used for analysis are available upon request.

Acknowledgements

The authors would like to acknowledge Apollo Hospitals Indraprastha, New Delhi, India for their support in providing the MRI samples for our study.

References

- 1 C. Palazzo, C. Nguyen, M. M. Lefevre-Colau, F. Rannou and S. Poiraudeau, Risk factors and burden of osteoarthritis, *Ann. Phys. Rehabil. Med.*, 2016, **59**(3), 134–138.
- 2 W. J. Scheuing, A. M. Reginato, M. Deeb and S. A. Kasman, *Best Pract. Res., Clin. Rheumatol.*, 2023, **37**, 101836.
- 3 M. V. Hurley, The role of muscle weakness in the pathogenesis of osteoarthritis, *Rheum. Dis. Clin. North. Am.*, 1999, **25**(2), 283–298.

- 4 N. Majumder, S. Seit, N. S. Bhabesh and S. Ghosh, An advanced bioconjugation strategy for covalent tethering of TGF β 3 with silk fibroin matrices and its implications in the chondrogenesis profile of human BMSCs and human chondrocytes: a paradigm shift in cartilage tissue engineering, *Adv. Healthcare Mater.*, 2024, **13**(10), 2303513.
- 5 J. Hermans, M. A. Koopmanschap, S. M. A. Bierma-Zeinstra, J. H. van Linge, J. A. N. Verhaar, M. Reijman and A. Burdorf, *Arthritis Care Res.*, 2012, **64**, 853–861.
- 6 R. Duncan, G. Peat, E. Thomas, E. M. Hay and P. Croft, *Ann. Rheum. Dis.*, 2011, **70**, 1944–1948.
- 7 A. D. Woolf, J. Erwin and L. March, *Best Pract. Res., Clin. Rheumatol.*, 2012, **26**, 183–224.
- 8 J. Chakraborty, J. Fernandez-Perez, M. T. Ghahfarokhi, K. A. van Kampen, T. ten Brink, J. Ramis, M. Kalogeropoulou, R. Cabassi, C. de J. Fernández, F. Albertini, C. Mota, S. Ghosh and L. Moroni, Development of 4D-bioprinted shape-morphing magnetic constructs for cartilage regeneration using a silk fibroin-gelatin bioink, *Cell Rep. Phys. Sci.*, 2024, **5**(3), 101819.
- 9 B. S. Dhinsa and A. B. Adesida, *Curr. Stem Cell Res. Ther.*, 2012, **7**, 143–148.
- 10 N. Majumder and S. Ghosh, *Adv. Funct. Mater.*, 2023, **33**, 2300651.
- 11 S. Saadeh, Z. M. Younossi, E. M. Remer, T. Gramlich, J. P. Ong, M. Hurley, K. D. Mullen, J. N. Cooper and M. J. Sheridan, *Gastroenterology*, 2002, **123**, 745–750.
- 12 A. Sambri, P. Spinnato, S. Tedeschi, E. Zamparini, M. Fiore, R. Zucchini, C. Giannini, E. Caldari, A. Cromb , P. Viale and M. De Paolis, *J. Pers. Med.*, 2021, **11**, 1317.
- 13 J. Chakraborty, N. Majumder, A. Sharma, S. Prasad and S. Ghosh, 3D bioprinted silk-reinforced Alginate-Gellan Gum constructs for cartilage regeneration, *Bioprinting*, 2022, **28**, e00232.
- 14 E. H. G. Oei, J. van Tiel, W. H. Robinson and G. E. Gold, *Arthritis Care Res.*, 2014, **66**, 1129–1141.
- 15 F. Aeffner, K. Wilson, N. T. Martin, J. C. Black, C. L. L. Hendriks, B. Bolon, D. G. Rudmann, R. Gianani, S. R. Koegler, J. Krueger and G. D. Young, *Arch. Pathol. Lab. Med.*, 2017, **141**, 1267–1275.
- 16 N. Majumder, C. Roy, L. Doenges, I. Martin, A. Barbero and S. Ghosh, *ACS Appl. Mater. Interfaces*, 2024, **16**, 9925–9943.
- 17 G. Iolascon, F. Gimigliano, A. Moretti, A. de Sire, A. Migliore, M. L. Brandi and P. Piscitelli, *Eur. Geriatr. Med.*, 2017, **8**, 383–396.
- 18 C. E. Quatman, C. M. Hettrich, L. C. Schmitt and K. P. Spindler, *Am. J. Sports Med.*, 2011, **39**, 1557–1568.
- 19 R. Kijowski, J. Fritz and C. M. Deniz, Deep learning applications in osteoarthritis imaging, *Skeletal Radiol.*, 2023, **52**(11), 2225–2238.
- 20 F. W. Roemer, A. Guermazi, S. Demehri, W. Wirth and R. Kijowski, *Osteoarthritis Cartilage*, 2022, **30**, 913–934.
- 21 F. W. Roemer, M. Jarraya, D. Hayashi, M. D. Crema, I. K. Haugen, D. J. Hunter and A. Guermazi, *Osteoarthritis Cartilage*, 2024, **32**, 460–472.
- 22 R. Lagier, Bone eburnation in rheumatic diseases: a guiding trace in today's radiological diagnosis and in paleopathology, *Clin. Rheumatol.*, 2006, **25**(2), 127–131.
- 23 D. J. Hunter, A. Guermazi, G. H. Lo, A. J. Grainger, P. G. Conaghan, R. M. Boudreau and F. W. Roemer, Corrigendum to Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score)[*Osteoarthritis and Cartilage* 2011; **19**: 990–1002], *Osteoarthritis Cartilage*, 2011, **19**(9), 1168.
- 24 M. Rafa , Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis, *ICT Express*, 2022, **8**(2), 183–188.
- 25 E. Calvo, I. Palacios, E. Delgado, J. Ruiz-Cabello, P. Hernandez, O. Sanchez-Pernaute, J. Egido and G. Herrero-Beaumont, High-resolution MRI detects cartilage swelling at the early stages of experimental osteoarthritis, *Osteoarthritis Cartilage*, 2001, **9**(5), 463–472.
- 26 X. Shi, Y. Mai, X. Fang, Z. Wang, S. Xue, H. Chen, Q. Dang, X. Wang, S. Tang, C. Ding and Z. Zhu, *Bone Rep.*, 2023, **18**, 101667.
- 27 D. Shen, G. Wu and H. I. Suk, Deep learning in medical image analysis, *Annu. Rev. Biomed. Eng.*, 2017, **19**(1), 221–248.
- 28 F. K. Nielsen, N. Egund, D. Peters and A. G. Jurik, *BMC Musculoskeletal Disord.*, 2014, **15**, 1–11.
- 29 F. Ambellan, A. Tack, M. Ehlke and S. Zachow, Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the Osteoarthritis Initiative, *Med. Image Anal.*, 2019, **52**, 109–118.
- 30 D. Muratovic, F. Cicuttini, A. Wluka, D. Findlay, Y. Wang, S. Otto, D. Taylor, J. Humphries, Y. Lee, A. Labrinidis, R. Williams and J. Kuliwaba, Bone marrow lesions detected by specific combination of MRI sequences are associated with severity of osteochondral degeneration, *Arthritis Res. Ther.*, 2016, **18**(1), 54.
- 31 L. Si, K. Xuan, J. Zhong, J. Huo, Y. Xing, J. Geng, Y. Hu, H. Zhang, Q. Wang and W. Yao, Knee cartilage thickness differs alongside ages: 3-T magnetic resonance research upon 2,481 subjects via deep learning, *Front. Med.*, 2021, **7**, 600049.
- 32 E. Calvo, I. Palacios, E. Delgado, O. Sanchez-Pernaute, R. Largo, J. Egido and G. Herrero-Beaumont, Histopathological correlation of cartilage swelling detected by magnetic resonance imaging in early experimental osteoarthritis, *Osteoarthritis Cartilage*, 2004, **12**(11), 878–886.
- 33 P. Omoumi, H. Babel, B. M. Jolles and J. Favre, Cartilage can be thicker in advanced osteoarthritic knees: a tridimensional quantitative analysis of cartilage thickness at posterior aspect of femoral condyles, *Br. J. Radiol.*, 2018, **91**(1087), 20170729.
- 34 O. Chamorro-Atalaya, J. Ar valo-Tuesta, D. Balarezo-Mares, A. Gonz lez-Pacheco, O. Mendoza-Le n, M. Quipuscoa-Silvestre, G. Tom s-Quispe and R. Suarez-Bazalar, *Int. J. Online Eng.*, 2023, **19**, 140–158.

- 35 T. Adugna, W. Xu and J. Fan, Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images, *Remote Sens.*, 2022, **14**(3), 574.
- 36 C. Martinez-Castillo, G. Astray, J. C. Mejuto and J. Simal-Gandara, Random forest, artificial neural network, and support vector machine models for honey classification, *eFood*, 2020, **1**(1), 69–76.
- 37 I. H. Sarker, *SN Comput. Sci.*, 2021, **2**, 1–21.
- 38 Q. Gazawy, S. Buyrukoğlu and Y. Yılmaz, Convolutional neural network for pothole detection in different road and weather conditions, *J. Comp. Electr. Electron. Eng. Sci.*, 2023, **1**(1), 1–4.
- 39 D. M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, *arXiv*, 2020, preprint, arXiv:2010.16061.
- 40 N. Aloysius and M. Geetha, A review on deep convolutional neural networks, in *2017 international conference on communication and signal processing (ICCSP)*, IEEE, 2017, April, pp. 0588–0592.
- 41 P. Molnar, T. P. Ahlstrom and I. Leden, Osteoarthritis and activity—an analysis of the relationship between eburnation, musculoskeletal stress markers (MSM) and age in two Neolithic hunter-gatherer populations from Gotland, Sweden, *Int. J. Osteoarchaeol.*, 2011, **21**(3), 283–291.
- 42 S. Ma, Y. Huang, X. Che and R. Gu, Faster RCNN-based detection of cervical spinal cord injury and disc degeneration, *J. Appl. Clin. Med. Phys.*, 2020, **21**(9), 235–243.
- 43 B. Liu, J. Luo and H. Huang, Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN, *Int. J. Comput. Assist. Radiol. Surg.*, 2020, **15**(3), 457–466.
- 44 S. Krishnapriya and Y. Karuna, Pre-trained deep learning models for brain MRI image classification, *Front. Hum. Neurosci.*, 2023, **17**, 1150120.
- 45 S. Tammina, Transfer learning using vgg-16 with deep convolutional neural network for classifying images, *Int. J. Sci. Res. Publ.*, 2019, **9**(10), 143–150.
- 46 X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo and J. Zhu, Pre-trained models: Past, present and future, *AI Open*, 2021, **2**, 225–250.
- 47 M. Gong, D. Wang, X. Zhao, H. Guo, D. Luo and M. Song, in *Seventh Symposium on Novel Photoelectronic Detection Technology and Applications*, SPIE, 2021, vol. 11763, pp. 821–828.
- 48 G. Vincent, R. Marchand, M. A. Mont, B. Harder, H. S. Salem, P. G. Conaghan, A. D. Brett and M. A. Bowes, Characterizing Osteophyte Formation in Knee Osteoarthritis: Application of Machine Learning Quantification of a Computerized Tomography Cohort: Implications for Treatment, *J. Arthroplasty*, 2024, **39**(11), 2692–2701.
- 49 J. Martel-Pelletier, P. Paiement and J. P. Pelletier, Magnetic resonance imaging assessments for knee segmentation and their use in combination with machine/deep learning as predictors of early osteoarthritis diagnosis and prognosis, *Ther. Adv. Musculoskeletal Dis.*, 2023, **15**, 1759720X231165560.
- 50 P. H. Conze, G. Andrade-Miranda, V. K. Singh, V. Jaouen and D. Visvikis, Current and emerging trends in medical image segmentation with deep learning, *IEEE Trans. Radiat. Plasma Med. Sci.*, 2023, **7**(6), 545–569.
- 51 J. Marques, R. Granlund, M. Lillholm, P. C. Pettersen and E. B. Dam, Automatic analysis of trabecular bone structure from knee MRI, *Comput. Biol. Med.*, 2012, **42**(7), 735–742.
- 52 W. Blanchard, The bone fragility and eburnation of rachitis, *J. Bone Jt. Surg.*, 1909, **2**(4), 616–625.
- 53 S. Nematzadeh, F. Kiani, M. Torkamanian-Afshar and N. Aydin, Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases, *Comput. Biol. Chem.*, 2022, **97**, 107619.
- 54 M. Sokolova and G. Lapalme, A systematic analysis of performance measures for classification tasks, *Inf. Process. Manage.*, 2009, **45**(4), 427–437.
- 55 J.-B. Schiratti, R. Dubois, P. Herent, D. Cahané, J. Dachary, T. Clozel, G. Wainrib, F. Keime-Guibert, A. Lalande, M. Pueyo, R. Guillier, C. Gabarroca and P. Moingeon, *Arthritis Res. Ther.*, 2021, **23**, 1–10.
- 56 M. Barbieri, A. A. Gatti and F. Kogan, Improving Accuracy and Reproducibility of Cartilage T 2 Mapping in the OAI Dataset Through Extended Phase Graph Modeling, *J. Magn. Reson. Imaging*, 2025, **61**(5), 2116–2127.
- 57 A. Kahveci, V. Alcan, M. Uçar, A. Gümüštepe, E. Bilgin, İ. Sunar and Ş. Ataman, *Clust. Comput.*, 2025, **28**, 1–11.
- 58 R. Zehbe, A. Haibel, H. Riesemeier, U. Gross, C. J. Kirkpatrick, H. Schubert and C. Brochhausen, Going beyond histology. Synchrotron micro-computed tomography as a methodology for biological tissue characterization: from tissue morphology to individual cells, *J. R. Soc., Interface*, 2010, **7**(42), 49–59.
- 59 L. Nagarajan, A. Khatri, A. Sudan, R. Vaishya and S. Ghosh, Deep learning augmented osteoarthritis grading standardization, *Tissue Eng., Part A*, 2024, **30**(19–20), 591–604.
- 60 A. S. Mohammed, A. A. Hasanaath, G. Latif and A. Bashar, Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed X-ray images, *Diagnostics*, 2023, **13**(8), 1380.