Medical & Biological Engineering & Computing
https://doi.org/10.1007/s11517-025-03405-y

ORIGINAL ARTICLE

Quantifying features from X-ray images to assess early stage knee osteoarthritis

Tanjina Helaly[1,2] · Tanvir R. Faisal[3] · Ahmed Suparno Bahar Moni[4] · Mahmuda Naznin[1]

Received: 12 March 2025 / Accepted: 20 June 2025

Abstract
Knee osteoarthritis (KOA) is a progressive degenerative joint disease and a leading cause of disability worldwide. Manual diagnosis of KOA from X-ray images is subjective and prone to inter- and intra-observer variability, making early detection challenging. While deep learning (DL)-based models offer automation, they often require large labeled datasets, lack interpretability, and do not provide quantitative feature measurements. Our study presents an automated KOA severity assessment system that integrates a pretrained DL model with image processing techniques to extract and quantify key KOA imaging biomarkers. The pipeline includes contrast limited adaptive histogram equalization (CLAHE) for contrast enhancement, DexiNed-based edge extraction, and thresholding for noise reduction. We design customized algorithms that automatically detect and quantify joint space narrowing (JSN) and osteophytes from the extracted edges. The proposed model quantitatively assesses JSN and finds the number of intercondylar osteophytes, contributing to severity classification. The system achieves accuracies of 88% for JSN detection, 80% for osteophyte identification, and 73% for KOA classification. Its key strength lies in eliminating the need for any expensive training process and, consequently, the dependency on labeled data except for validation. Additionally, it provides quantitative data that can support classification in other OA grading frameworks.

Keywords Knee osteoarthritis · X-ray image · Quantification · Joint space narrowing · Intercondylar osteophyte

1 Introduction

Osteoarthritis (OA) is a progressive degenerative joint disease that primarily affects articulating joints, with the knee, hip, shoulder, and hand being the most commonly affected areas [1, 2]. Approximately 22% of the world's population over 40 years old is affected by OA [3]. Among all articulating joints, *knee osteoarthritis* (KOA) has the highest prevalence [1, 2]. A comprehensive population-based survey across 91 countries, conducted by the Global Burden of Disease (GBD) 2021 Osteoarthritis Collaborators [4], revealed that 7.6% of the global population experienced OA in 2020, representing a 132.2% increase since 1990. Furthermore, KOA is projected to increase by an additional 74.9% by 2050. Given the irreversible nature of KOA, early detection is crucial to slow disease progression through lifestyle modifications and conservative treatments. Without early intervention, patients in the advanced stages of KOA have no alternative other than total knee replacement (TKR) surgery.

The identification of visible abnormalities through various imaging modalities serves as a primary diagnostic tool for KOA. X-ray imaging is widely used by both orthopedic surgeons and primary care physicians for the diagnosis of OA due to its low radiation exposure and cost-effectiveness. While magnetic resonance imaging (MRI) [5–7], computed tomography (CT) [8, 9], and acoustic analysis [10, 11] pro-

✉ Tanvir R. Faisal
tanvir.faisal@louisiana.edu

Tanjina Helaly
tanjina@uap-bd.edu

Ahmed Suparno Bahar Moni
AhmedSuparno.BaharMoni@utoledo.edu

Mahmuda Naznin
mahmudanaznin@cse.buet.ac.bd

1 Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

2 Department of Computer Science and Engineering, University of Asia Pacific, Dhaka, Bangladesh

3 Department of Mechanical Engineering, University of Louisiana at Lafayette, Lafayette, LA, USA

4 Department of Orthopedic Surgery, College of Medicine and Life Sciences, University of Toledo, Toledo, OH, USA

Published online: 05 July 2025

vide additional diagnostic insights, X-ray imaging remains the gold standard. The Kellgren-Lawrence (KL) grading system is commonly used to assess the severity of OA, classifying the disease into five stages (KL 0-4) based on radiographic biomarkers and clinical features (Fig. 1). Radiographic indicators include *joint space narrowing* (JSN), presence of *osteophytes*, *subchondral sclerosis, subchondral cysts*, and *bone deformity*, while clinical features commonly encompass *pain level*, *morning stiffness*, and *age* [12–16].

Early KOA detection presents significant challenges, as initial manifestations—such as mild JSN and osteophytes in the intercondylar region [17–19]—are often subtle and difficult to detect visually from X-ray images. This leads to intra- and inter-observer variability [20, 21], increasing the risk of delayed diagnosis and disease progression. Furthermore, the qualitative nature of the assessment criteria of the grading system makes diagnosis susceptible to subjective interpretation and inconsistency, particularly in the identification of early-stage KOA. Given the limitations of the KL grading system, automated detection of KOA using X-ray images has gained significant attention in recent research. The detection process consists of two primary steps: first, identifying the region of interest (ROI) containing key KOA-related features, and second, classifying KOA severity based on these extracted features. Current research in automated radiological diagnosis of KOA encompasses three main methodological approaches: (i) image processing-based detection models [22], (ii) machine learning/deep learning (ML/DL)-based detection models [23–27], and (iii) hybrid detection models that combine image processing, statistical, and ML/DL algorithms. In addition, researchers have explored manual detection methods using computer-aided techniques [28–32]. Despite their widespread adoption, the approaches based on ML and DL face several significant challenges. These include the requirement for large volumes of labeled data, which demands substantial resources and financial investment. Another notable limitation is their insufficient emphasis on intercondylar osteophytes, which serve as an important early indicator of KOA. Furthermore, these approaches often fail to extract quantifiable features, which limits their interpretability.

To address the current limitation in KOA assessment, this study introduces an automated hybrid system to quantify the key radiological features, specifically the percentage of JSN and the quantity of osteophytes in the intercondylar region—two critical indicators. In this proposed framework, the dense extreme inception network for edge detection (DexiNed) [33] edge detection model is utilized in a *zero-shot inference mode* to extract high-resolution edge maps from X-ray images without any additional task-specific training on knee X-ray data. The edge-detected output is later processed through an image processing pipeline designed to facilitate a detailed and interpretable quantification of the severity of KOA. This novel quantitative framework advances the field in two significant ways: it enhances diagnostic consistency through objective measurements, and it improves the detection of early-stage KOA by providing reproducible assessment criteria. The system directly addresses the pressing clinical need for objective and standardized evaluation methods in the diagnosis of KOA.

The key contributions of this research can be listed as follows:

- **Quantitative joint space narrowing analysis:** This study presents a reliable computational framework that quantifies JSN in the tibiofemoral joint with high accuracy. The system integrates error-tolerant processing to address edge discontinuities and noise, enabling consis-

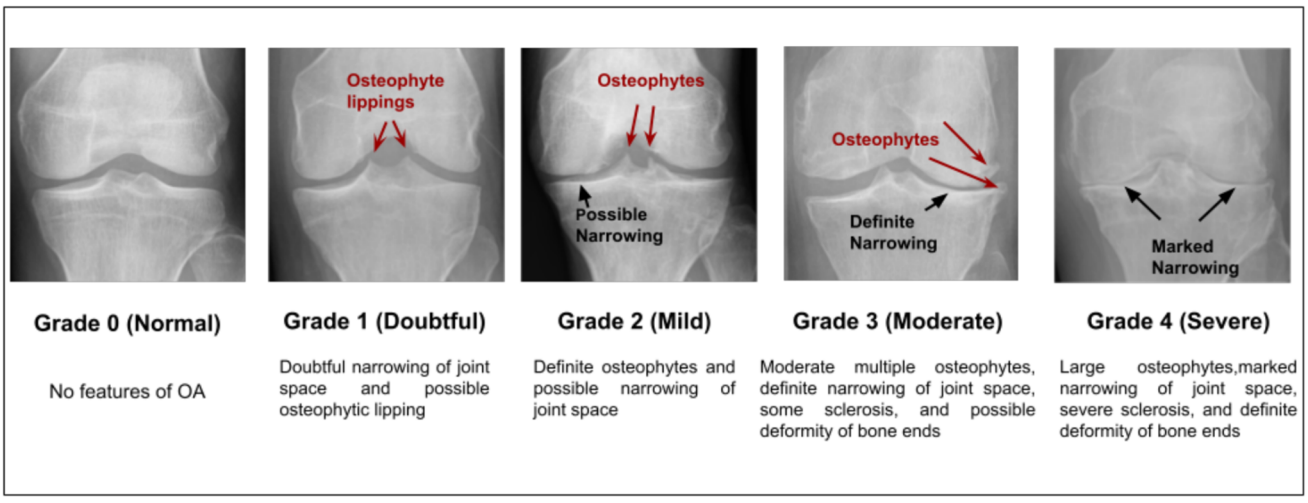


Fig. 1 KL severity grading on sample X-ray images with manually identified osteophytes and different levels of JSN commensurate with the grading

tent performance across variable radiographic quality, ensuring practical clinical utility.

- **Automated osteophyte detection in the intercondylar region:** A targeted edge-based pipeline is developed to detect and quantify osteophytes in the intercondylar region, a key early radiographic marker of KOA. The approach accurately isolates osteophytic growth from adjacent bone structures, offering precise, objective insight into early disease progression.
- **Severity classification using quantitative radiographic biomarkers:** An enhanced grading framework is proposed combining quantitative measurements of JSN and osteophytes to generate objective KOA severity scores. By establishing standardized, data-driven criteria, the system minimizes subjectivity and inter-observer variability, thereby improving diagnostic reliability and enabling consistent tracking of disease progression over time.
- **Cost-effective zero-shot KOA diagnosis:** Unlike conventional ML/DL approaches that depend heavily on manually labeled datasets, our system avoids the need for expensive training and hence expert annotation by leveraging a pretrained deep learning model for edge detection in a zero-shot inference setting. Expert annotations are reserved exclusively for validation purposes. Coupled with classical image processing techniques, this hybrid strategy offers a scalable, cost-effective solution for automated KOA diagnosis, maintaining clinical-level accuracy while reducing implementation overhead.

2 Related work

A critical analysis of existing KOA detection approaches reveals several notable methodological limitations. Mahrukh Saleem et al. [22] have developed a multi-phase image processing system utilizing a histogram of oriented gradients (HOG)-based template matching algorithm for ROI identification, followed by Canny edge detection for tibiofemoral joint edge estimation. Their method calculates the joint space in both medial and lateral compartments to determine the presence of KOA. However, this approach presents two significant limitations: it does not assess disease severity levels, and it overlooks the crucial role of osteophytes in disease progression. Other researchers [24, 34–36] have explored hybrid approaches combining image processing with ML and DL models. While these studies contribute to the field, they primarily focus on traditional qualitative grading detection without implementing quantitative measurements. This limitation reduces the objective assessment capabilities necessary for precise disease monitoring and progression tracking.

The majority of KOA studies employ deep learning for both detection and classification. Convolutional neural networks (CNNs) [35, 37] and transfer learning models, including CNN Inception Net V2 [34, 38], ResNet [26, 38, 39], VGG [26, 39], DenseNet [26, 38], Xception [38, 40], and InceptionV3 [26], along with Visual Transfer Model [23], have been commonly employed for classification tasks. Neural architecture search (NAS) [25] has recently gained prominence as an effective tool for developing optimized model architectures. DL-based approaches are employed for ROI detection and feature extraction prior to employing DL-based classification models. For ROI detection, methods like faster R-CNN [37] and local center of mass (LCM) [34] are utilized. Feature extraction is performed using either statistical histogram-based approaches [35] or DL models such as ResNet [23] and EfficientV2 [25].

Recent research has demonstrated the effectiveness of ensemble learning techniques in advancing KOA classification accuracy. Tariq et al. [41] have developed a sophisticated ensemble model by integrating predictions from multiple fine-tuned base classifiers, including DenseNet 161, DenseNet 121, ResNet-34, and VGG. Their architecture employs a fully connected final layer, optimized using the Adam algorithm, to aggregate outputs from these base models effectively. Building on this approach, Pi et al. [42] have proposed an innovative ensemble network that combines different versions of DenseNet, EfficientNet, ResNet, and ShuffleNet. Further advancing the field, Phan et al. [43] have introduced a hybrid model that integrates DenseNet and EfficientNetV2 architectures, incorporating a fully connected classification layer. The system employs YOLOv5 for ROI detection and has been successfully implemented in a dedicated application, DIKOApp. To validate and interpret these DL models, researchers have employed various explainable AI visualization tools. Notable among these are EigenCAM [41] and GradCAM [42, 44], which provide insights into the model's decision-making process by highlighting the regions most influential in classification decisions. These DL and ensemble approaches face two major limitations. First, they require extensive labeled datasets, which are costly and time-consuming to generate, particularly given the scarcity of orthopedic experts. Second, these methods fail to provide quantitative measurements of specific radiological features, limiting their clinical utility for objective disease assessment and progression monitoring.

Some researchers have also explored hybrid approaches combining statistical, morphological, and DL models for KOA assessment. For feature extraction, researchers have employed diverse techniques: complex network-based approaches [45], CNNs [46], combined CNN-HOG-LBP models [24], integrated LBP-DarkNet-AlexNet systems [36], multi-feature approaches incorporating LBP and statistical measures [44], and UNet-based segmentation with geomet-

ric and radiomic feature extraction [47]. For classification tasks, researchers have employed various machine learning algorithms. Support vector machine (SVM), random forest (RF), K-nearest neighbors (KNN), linear discriminant analysis (LDA), gradient boosting (GB), and XGBoost (XGB) are prominently featured across multiple studies. Specific implementations include combinations of SVM, KNN, and LDA [45]; SVM, RF, and KNN [24]; SVM and KNN [36]; standalone SVM [46]; RF, KNN, GB, and XGB [44]; and hierarchical classification models using logistic regression, RF, SVM, and XGBoost [47]. Despite their sophistication, these hybrid approaches inherit the limitations of DL models: dependence on extensive labeled datasets and the lack of quantitative assessment of radiological features. While Yoon et al. [48] have developed a DL-based software that provides quantitative measurements, including JSN percentage and presence of marginal osteophyte, these measurements are not integrated into the severity classification system. This study addresses these limitations by proposing an efficient hybrid model that not only quantifies JSN and osteophytes but also correlates these measurements with KOA severity classifications.

3 Proposed approach

This section presents a comprehensive description of the proposed framework. The complete system architecture and workflow are illustrated in Fig. 2.

3.1 Dataset

Widely recognized "Knee Osteoarthritis Severity Grading Dataset" [49] from Osteoarthritis Initiatives (OAI), which provides systematically annotated samples based on KL grading, has been utilized in this study. We have analyzed a carefully selected subset of this dataset, developing a detailed categorization system based on JSN severity and osteophyte distribution patterns. The JSN classification system comprises five distinct categories: JSN 0 (No JSN), JSN 1 (<25% JSN), JSN 2 (25% to 50% JSN), JSN 3 (51% to 75% JSN), and JSN 4 (>75% to 100% JSN). To establish a reference baseline, benchmark gaps have been computed using the healthy samples from the dataset. Precise pixel-wise distance measurements are obtained using an open-source tool ImageJ [50] that captures both pixel coordinates and intensity values from the radiographic images.

The assessment of intercondylar osteophytes has followed a rigorous two-stage validation process. Initial annotations are performed by a trained evaluator who documents both the presence and quantity of osteophytes in the intercondylar region. These annotations undergo subsequent review and refinement by an experienced orthopedic surgeon to ensure diagnostic accuracy. This iterative review process is implemented specifically to minimize the risk of misclassification error and enhance the reliability of our annotated dataset. To be noted that the annotated data set is used solely to validate the performance of the model and is not involved in the training process.

3.2 Pre-processing: contrast enhancement with CLAHE

Image preprocessing plays a critical role in the proposed framework, particularly given the inherent complexity of X-ray images. Contrast-limited adaptive histogram equalization (CLAHE) [51] is the primary preprocessing technique implemented to enhance image contrast, which significantly impacts the accuracy of subsequent edge detection processes. CLAHE operates by partitioning the image into small tiles and independently enhancing the contrast within each tile using histogram equalization. The technique employs a clip limit parameter to prevent over-amplification of noise, maintaining image integrity. A systematic evaluation was conducted to determine the optimal clip limit value by analyzing image histograms and assessing the quality of resulting edge-detected images across various parameter settings. This analysis established that a clip limit of 2.0 provides the optimal balance between feature preservation and noise reduction. This setting effectively maintains the visibility of important anatomical structures, particularly in the *intercondylar* region, while minimizing image noise. The default tile grid size of 8×8 was used in this study.

3.3 Edge detection

Following image preprocessing, the framework advances to joint boundary extraction, a crucial step for both the quantitative measurement of JSN and the identification of osteophytes. For this purpose, DexiNed [33], the high-capacity pretrained deep neural network, is utilized to perform precise edge delineation on the enhanced X-ray images. The model, originally trained on the Barcena Images for Perceptual Edge Detection (BIPED) dataset, is deployed without any further training or fine-tuning, thus bypassing the need for task-specific data partitioning, resource-intensive annotation, and retraining.

Unlike object detection tasks, which require training on object-specific semantic features (e.g., texture, shape, context) and typically involve bounding boxes or region proposals, edge detection focuses solely on generic low-level visual cues, primarily local intensity gradients and discontinuities. These features are domain-independent, making edge detection models, such as DexiNed, highly transferable across image domains, including medical imaging, without the need for contextual learning of target-specific

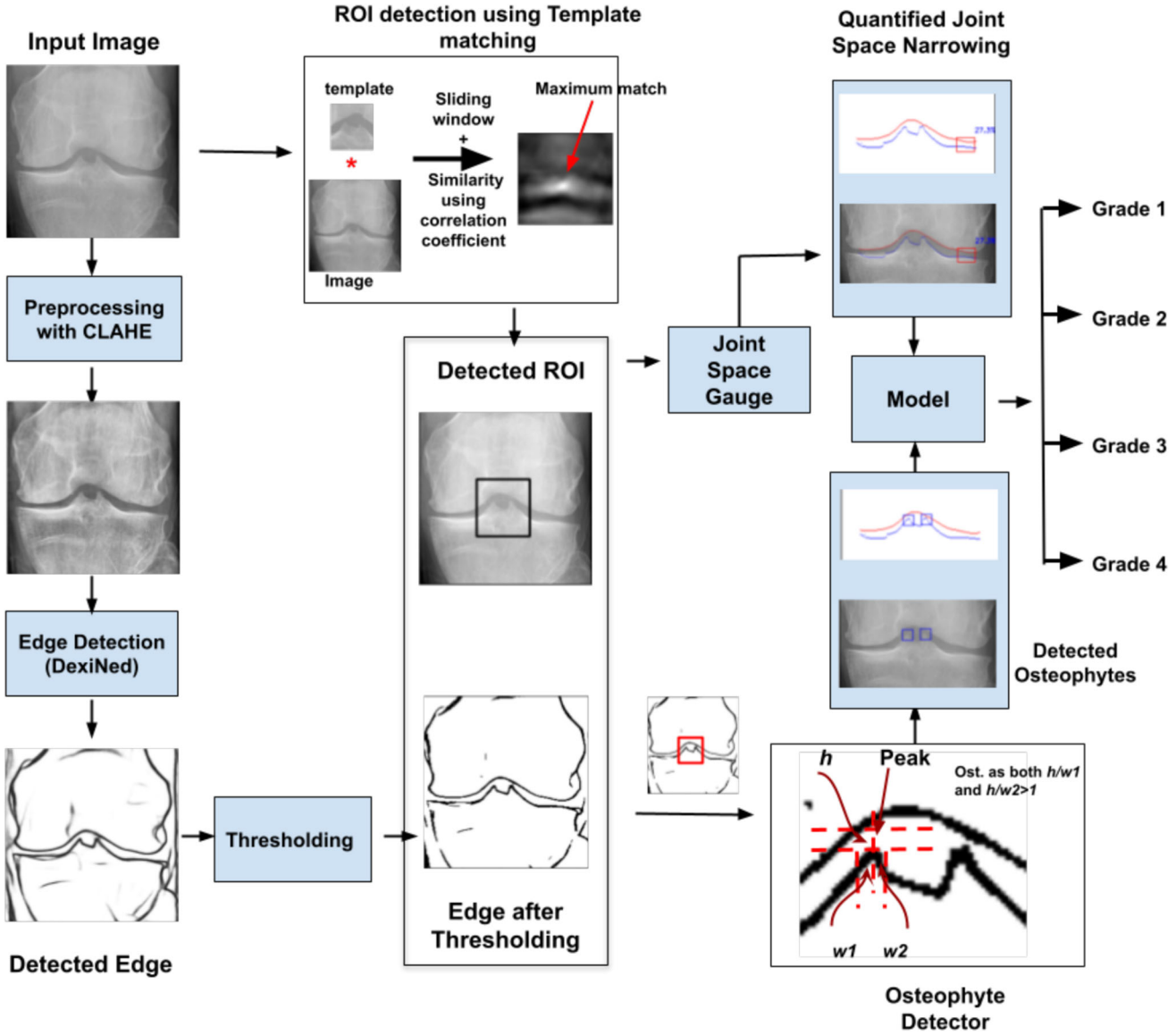


Fig. 2 Workflow of the proposed system to assess JSN and intercondyler osteophytes

attributes. Moreover, DexiNed's inception-based multiscale feature representation further enhances its ability to capture the complex, irregular, and fine boundaries often present in knee joint radiographs, thereby supporting both structural feature extraction and downstream quantitative analysis.

DexiNed, inspired by both holistically-nested edge detection (HED) [52] and Xception networks, is designed to produce a thin edge map. Figure 3 shows the architecture of the DexiNed, which comprises a stack of learned filters that take an image as input and predict an edge map with the same resolution. DexiNed comprises two sub-networks: the dense extreme inception network (Dexi) and the up-sampling block (UB).

The Dexi sub-network, inspired by the Xception network, has been fed with the RGB images and produces feature maps. The output of the previous Dexi block is the input of the next, and the Dexi blocks have been connected between them through 1x1 convolutional blocks. The UB sub-network consists of conditional stacked sub-blocks, each containing two layers: one convolutional and one deconvolutional. This sub-network receives feature maps from each block of Dexi and produces intermediate edge maps. These edge maps are then concatenated and passed through a final stack of learned filters. The resulting network produces a thin, fused edge map, reducing the chances of missing edges in the deeper layers.

DexiNed was selected for edge detection following a comprehensive evaluation of various methods, including traditional image processing techniques and machine learning-based approaches. The traditional methods examined include

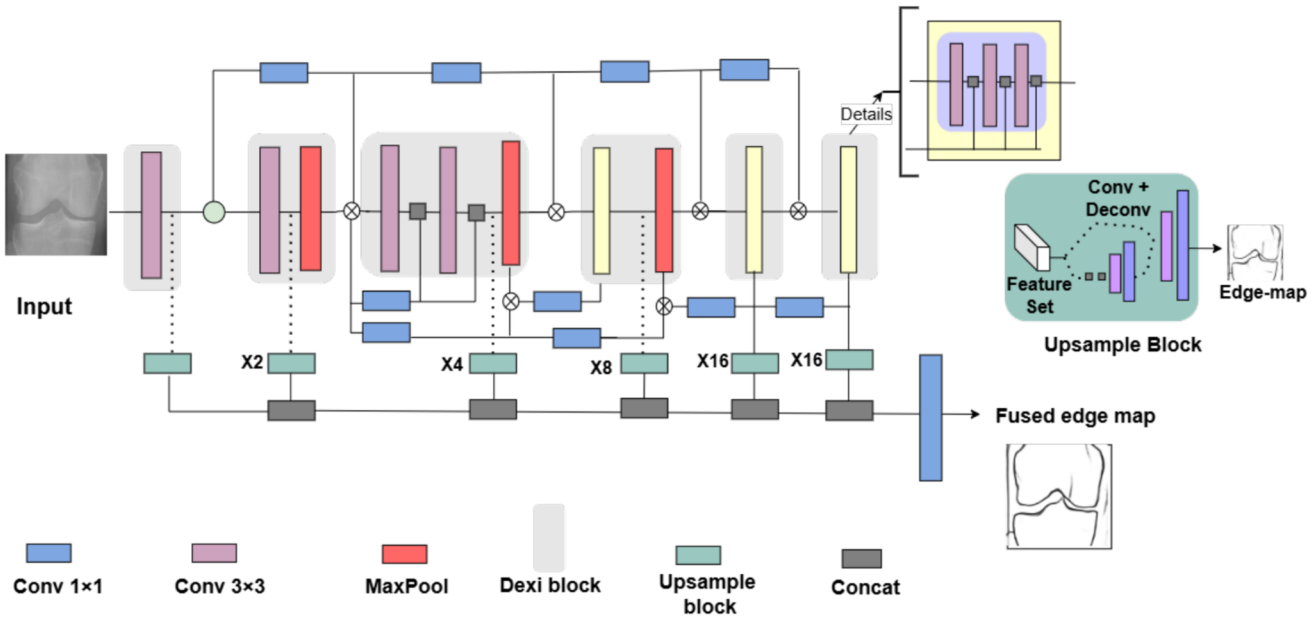


Fig. 3 A generalized DexiNed architecture for extracting the edge of the tibiofemoral joint from X-ray image

Canny, Sobel, and Prewitt edge detection techniques [53, 54]. These are evaluated alongside advanced ML models such as HED, DexiNed, and lightweight dense CNN (LDC) [55]. The performance of each technique was quantitatively assessed using the following mathematical equation:

$$performance = 1 - p \times no.\ of\ missed\ edges - q \times no.\ of\ noisy\ edges \quad (1)$$

where p and q are the weight coefficients. Figure 4 presents some sample outputs from a subset of evaluated methods.

To evaluate the edge detection performance, the joint boundary was divided into three anatomically defined regions: medial, intercondylar, and lateral. Each region contains two key structural edges—one corresponding to the femoral surface and the other to the tibial surface—resulting in six clinically significant edges for evaluation. Within the proposed metric, missed edges are defined as any of these six anatomical edges that remain undetected, while noisy edges denote spurious detections occurring within the vicinity of these regions. The performance equation incorporates weighted coefficients, setting coefficients for missed edges (p) greater than that for noisy edges (q), reflecting the greater clinical relevance of accurately detecting anatomical boundaries. Based on comprehensive testing across X-ray datasets representing varying stages KOA, the combination of DexiNed and CLAHE-based preprocessing consistently delivered the most accurate and robust results, establishing it as an optimal configuration for this application.

3.4 Thresholding

Edge localization accuracy requires effective noise reduction and binary image conversion. To achieve this, thresholding is applied to the DexiNed-generated edge maps. The optimal threshold value is determined through a combination of histogram analysis, percentile-based evaluation, and empirical testing. Since the edge maps predominantly contain white areas with thin black edge lines and exhibit histogram peaks near a pixel intensity of 255, it was statistically observed that black pixels typically fall below the 25th percentile in intensity. Through qualitative evaluation of multiple threshold values below this percentile, a threshold of 100 has been identified as optimal, yielding clear edge delineation with minimal background noise.

3.5 Feature extraction and quantification

3.5.1 ROI detection using template matching

ROI selection plays a critical role in optimizing computational efficiency and minimizing false detections. In the proposed framework, analysis is confined to the joint area, specifically encompassing a few millimeters above and below the articular surface. For automated ROI identification, we have implemented the template matching algorithm from the OpenCV [56] library. This algorithm operates by sliding a predefined template across the image and computing correlation coefficients between the template and local image patches. The resulting correlation map identifies regions of high similarity, effectively localizing the joint area. This pre-

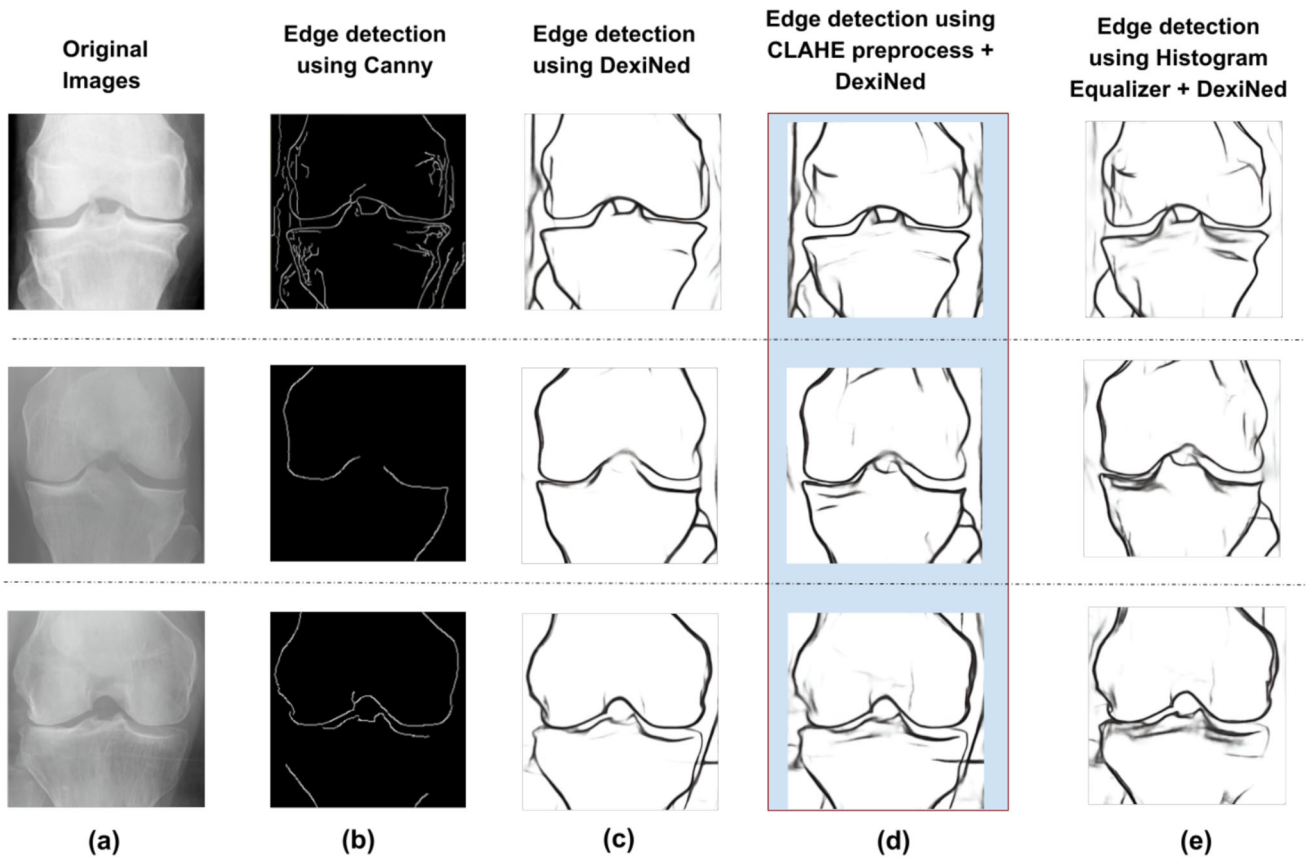


Fig. 4 Comparison among the outputs of different edge detection methods. **a** The original images, **b** edges of the images are generated using *Canny* algorithm—it completely misses some edges and sometimes has more noisy edges, **c** edges of images are generated using DexiNed pretrained edge detection model—it produces better edges in medial and lateral side; however, it sporadically misses the edges in inter-condylar region which are essential to identify inter-condylar ostephytes, **d** edges of images are generated using CLAHE pre-processing followed by DexiNed—this process identifies the inter-condylar edges better; however, it introduces noise. The noise gets minimized in the thresholding step, and **e** edges of the images are generated using *histogram equalizer* pre-processing, followed by DexiNed—it introduces more noise. Among these, the combination of CLAHE preprocessing with DexiNed (highlighted in the boxed region in **d** has produced the most accurate and clinically meaningful images, and therefore, **d** has been selected as the preferred method for this study

cisely identified ROI serves as the foundation for subsequent analyses, including joint space width (JSW) calculation and intercondylar osteophyte detection.

3.5.2 Detection of joint space narrowing

The calculation of joint space and the identification of narrowing requires precise determination of edge coordinates. While this process is relatively straightforward in ideal conditions with clear edge definition, X-ray images frequently present challenges due to haziness, resulting in either spurious or missing edges. These imaging artifacts necessitate a sophisticated approach to accurately identify the femoral and tibial edges. The detailed methodology for edge coordinate extraction is presented in Algorithm 1, while the complete process for JSN detection is thoroughly documented in Algorithm 5.

Following ROI detection, the joint area is partitioned into eight specific regions (lines 1–9): three in the lateral compartment, three in the medial compartment, and two in the intercondylar region, as shown in Fig. 5. This regional subdivision is based on characteristic edge trajectory patterns. Initial detection of femur and tibia edge points is performed separately for the left and right sides using the DetectInitialEdges() function (line 12 and detailed in Algorithm 2). Subsequently, a column-wise scan is initiated from the initial edge points toward the image center (lines 14–23) to identify and trace edge points. Within each column, potential edge candidates are identified by scanning vertically within the ROI boundaries and linking them to the existing femur or tibia edges. If multiple edges are connected (lines 17–

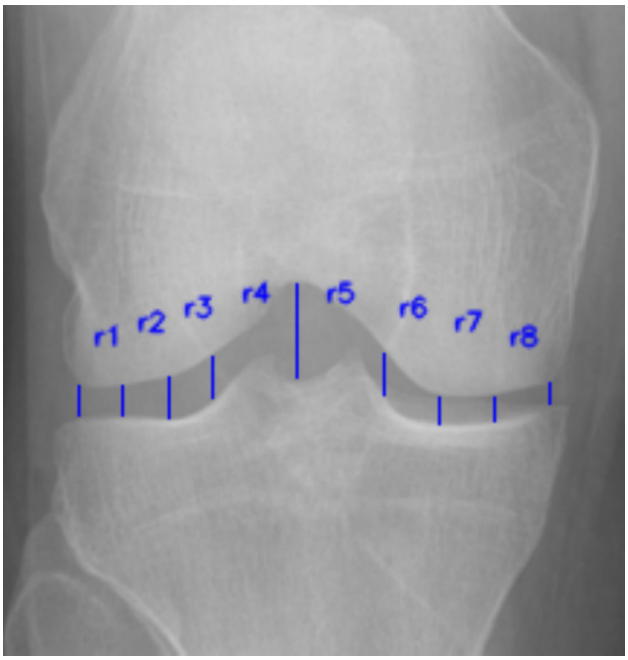


Fig. 5 Regional division of the knee joint for JSN assessment. The knee joint is segmented into specific anatomical regions to facilitate localized evaluation. The intercondylar region, identified as the ROI using *template matching*, is divided into two equal subregions. The lateral and medial compartments, located outside the ROI, are each further divided into three equal regions, resulting in a total of eight regions across the joint. Each region exhibits an approximate edge slope pattern: r1 and r8: slope decreases inward toward the joint center. r2 and r7: slope remains approximately constant. r3 and r6: slope increases inward

18), the most valid one is selected based on region-specific slope patterns, horizontal span, and proximity to the expected edge gap as illustrated in Algorithm 3. In the event of edge discontinuity (lines 19–20), the algorithm uses surrounding edge information to infer missing segments through interpolation, leveraging slope consistency and vertical alignment, as described in Algorithm 4. This process begins by identifying a valid forward point, $point_{found} = (c_p, r_p)$, which must meet specific conditions. If the missing segment falls within a horizontal region (r2 or r7), the algorithm fills it using the same row value (r_p) across intermediate columns. Otherwise, interpolation is applied based on the slope between the last known point and the identified point. To ensure valid prediction, the missing segment must be confined to a single region, with adjacent regions (e.g., r1 and r3 for r2) providing contextual support. Each detected or interpolated femur–tibia coordinate pair is appended to xRayEdgeDataList (line 24), which is returned as the final output (line 26). The performance of the algorithm in effectively recognizing edges and reconstructing missing segments amidst noise is demonstrated in Figs. 6 and 7.

After successful edge identification, joint space widths are calculated using the extracted edge coordinates. Areas

Algorithm 1 Extracting edge information (coordinates).

Require: Edged image, Benchmark gap, coordinates of ROI (output of template matching)
Ensure: $xRayEdgeDataList$ - Extracted edge data
1: **Step 1:** Divide the ROI into 8 sub-regions based on the slope pattern of edge (shown in Fig. 5): ▷ These regions have distinct slope patterns: ($r1, r8$: downward), ($r2, r7$: horizontal), ($r3, r6$: upward)
2: $[r_{top}, c_{top}]$: : Coordinate of top left corner of ROI
3: $[r_{bottom}, c_{bottom}]$: Coordinate of bottom right corner of ROI.
4: $w \Leftarrow imagewidth$
5: $left_{bound} \Leftarrow w/8$
6: $right_{bound} \Leftarrow 7 * w/8$
7: Divide center ROI into: $r4$: $[c_{top}$ to $mid]$ and $r5$: $[mid$ to $c_{bottom}]$ where $mid = (c_{top} + c_{bottom})/2$
8: Divide left side ($left_{bound}$ to c_{top}) into Regions $r1$, $r2$, and $r3$ of equal width.
9: Divide right side ($right_{bound}$ to c_{bottom}) into Regions $r6$, $r7$, and $r8$ of equal width
10: ▷ The following steps are repeated twice: 1) For the left half and 2) For the right half. The Scanning is done from outside to inside i.e, for the left half, scan from left to right, for the right half, right to left.
11: **Step 2:** Detect the initial femur and tibia edge point and store to $femurEdge, tibiaEdge$
12: DetectInitialEdges() - Illustrated in Algorithm 2
13: **Step 3:** Scan column-wise and identify the femur and tibia points and store to $femurEdge, tibiaEdge$
14: **while** Scan column-wise from the initial point (femur/tibia) to the mid of the image **do**
15: Scan row-wise (r_{top} to r_{bottom}) and Identify all possible edge points (connected black pixels)
16: Identify the edge points that are connected to $femurEdge$ and $tibiaEdge$
17: **if** Multiple edges are connected **then**
18: IdentifyCorrectEdge() - Illustrated in Algorithm 3
19: **else if** Either $femurEdge$ or $tibiaEdge$ is missing **then**
20: PredictMissingEdgeSegment() - Illustrated in Algorithm 4
21: **else**
22: Assign the respective connected point to $femurEdge, tibiaEdge$
23: **end if**
24: Append $[c, femurEdge, tibiaEdge]$ to $xRayEdgeDataList$
25: **end while**
26: **return** $xRayEdgeDataList$

exhibiting widths below the established benchmark gap are classified as regions of JSN and annotated accordingly. The benchmark gaps for both lateral and medial compartments are established through the analysis of healthy knee radiographs. Our measurements confirm the anatomically expected pattern where the lateral compartment demonstrates a wider gap compared to the medial compartment. The quantification of JSN is computed using Eq. 2.

$$JSN = \begin{cases} 100 \times \frac{JSW_{min} - JSW}{JSW_{avg}} & if\ JSW < JSW_{min} \\ 0, & otherwise \end{cases} \quad (2)$$

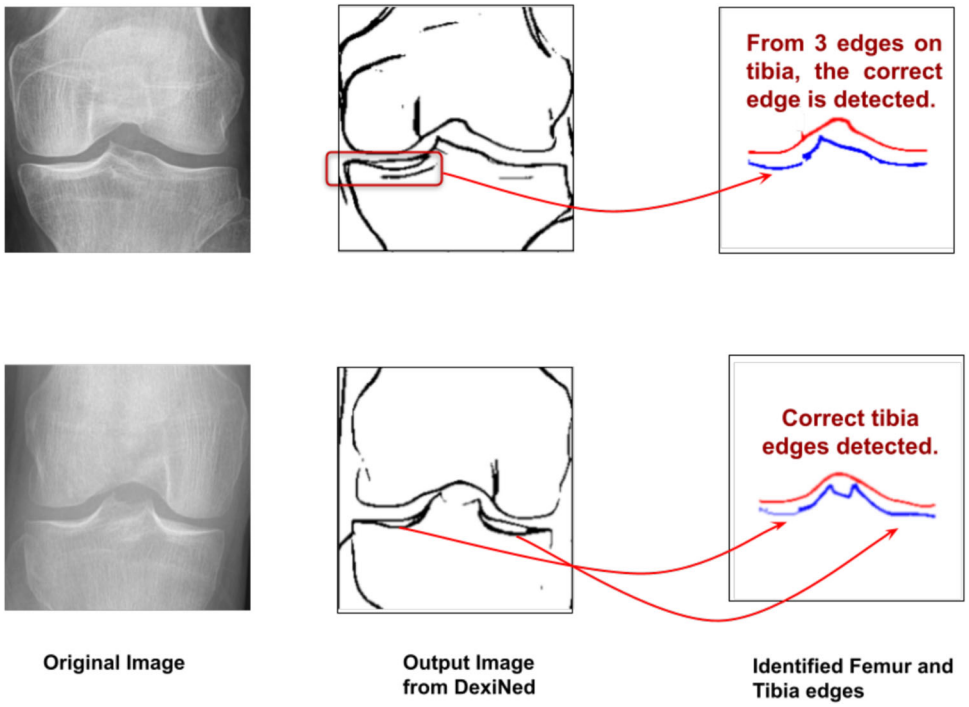


Fig. 6 Edge identification process from noisy image

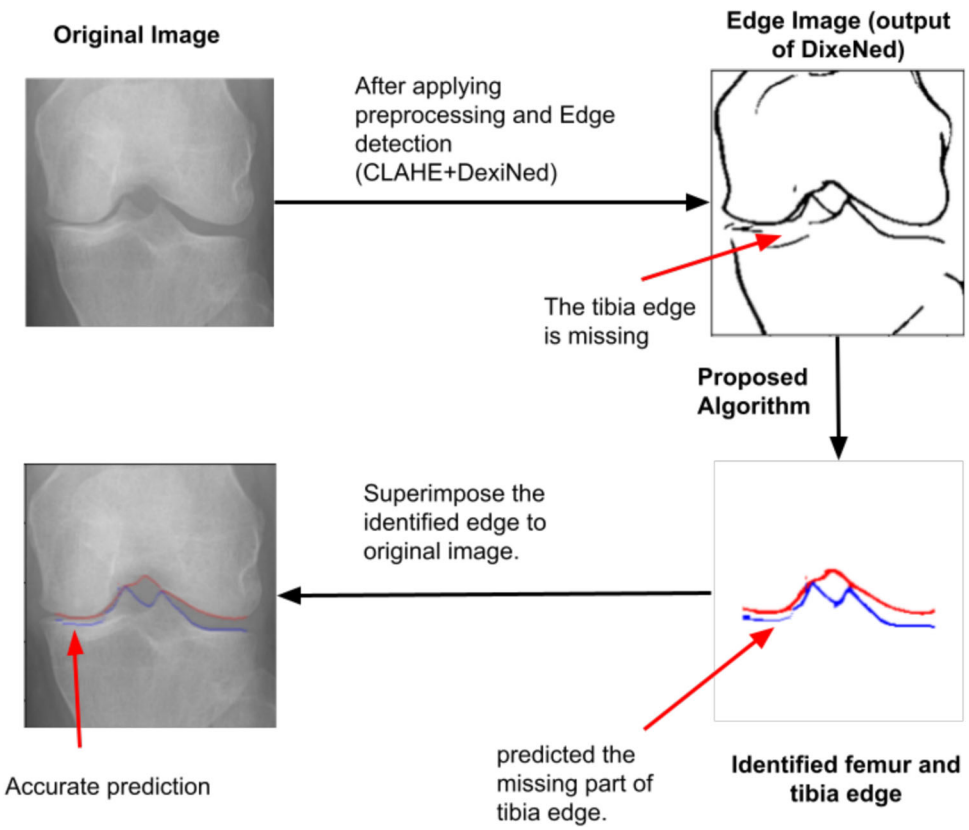


Fig. 7 Missing edge prediction

Algorithm 2 DetectInitialEdges(): Detecting the initial Femur and Tibia edge points (coordinates).

Require: Edged image, Benchmark gap ($gap_{benchmark}$), coordinates of ROI ([r_{top}, c_{top}] and [r_{bottom}, c_{bottom}]), $left_{bound}$, $Right_{bound}$ regions: $r1$ - $r8$, isLeftHalf (value to represent if scanning in left half or right)
Ensure: First point on the femur and tibia edge
For Left Half: Start at column $c = left_{bound}$, scan from r_{top} to r_{bottom}:
For Right Half: Start at column $c = Right_{bound}$, scan from r_{top} to r_{bottom}:
while points are not identified and c is in $r1$ (For Left half) or $r8$ (for Right half) **do**
    Locate black pixels
    for $item$ in black pixels **do**
        Find horizontal neighbor count -connected black points
        Identify slope pattern - downward, horizontal, upward.
    end for
    if 2 valid points found with vertical span >= half of a region and slope pattern matches (downward) and the vertical gap between the points < 2*$gap_{benchmark}$ **then**
        Assign the first point as the femur and the last as the tibia.
    else if More than 2 points found **then**
        Check the location of the points.
        if points are too close to the boundary (r_{top} and r_{bottom}) **then**
            probability of being a femur and tibia edge is low.
        end if
        Based on location, gap between 2 points, identify 2 points as femur and tibia.
        if Unable to identify 2 such points **then**
            Go to next c
        end if
    else
        Go to next c
    end if
end while

Algorithm 3 IdentifyCorrectEdge(): Identify the next Femur and Tibia edge points (coordinates) from possible points - based on the slope pattern of that region.

Require: Edged image, connected edge poinss($points_{conn}$). Benchmark gap ($gap_{benchmark}$), regions: $r1$ - $r8$, isLeftHalf (value to represent if scanning in left half or right)
Ensure: Next point connected to the femur and tibia edge
$gap_{allowed} \Leftarrow gap_{benchmark}/2$
$span_{allowed} \Leftarrow half\ span\ of\ any\ medial\ or\ lateral\ region$
for each point, $point_{cur}$ in $points_{conn}$ **do**
    Identify the horizontal connected neighbours and the distance to the last neighbor within neighboring region, $span_{hor}$
    if sporadic edge ($span_{hor} < span_{allowed}$) or outside of $gap_{allowed}$ gap or slope pattern doesn't match for that region **then**
        mark as noise
    else if $span_{hor} >= span_{allowed}$ and within $gap_{benchmark}$ gap and slope pattern matches **then**
        if the new point is farther away from the previous tibia edge **then**
            Reassign the new point as part of the tibia edge
            Update the previously recorded tibia edge with the connected point of this new one
        end if
    end if
end for

Algorithm 4 PredictMissingEdgeSegment(): Predict the missing part based on the slope pattern of that region, and the adjacent edge points (Precondition: The missing part should be confined to a single region).

Require: Edged image, last found edge, Benchmark gap ($gap_{benchmark}$), regions: $r1$ - $r8$, isLeftHalf (value to represent if scanning in left half or right)
Ensure: Predicted edge
Use (c-1) edge value as reference edge information, where c is the column where the missing edge starts
Scan forward to find the next valid point that satisfies conditions: i) vertical distance of the point is within the range of 2*$gap_{benchmark}$ from femur/tibia, and ii) the slope pattern matches with the region.
if found such a point, $Point_{found}$=(c_p, r_p) **then**
    if Missing edge is in region r2, r7 **then**:
        Populate the missing edges from column, $c_i = c$ to $c_p - 1$ with (c_i, r_p)
    else
        Calculate the slope between the last and the found point.
        Interpolate the missing part using the calculated slope.
    end if
end if

Algorithm 5 Joint space narrowing detection.

Require: xRayEdgeData, $minGap$(minimum benchmark gap)
Ensure: JSN Data: $jsnData$
$jsnData$=[]
for $edgeInfo$ in $xRayEdgeData$ **do**
    $JSW \Leftarrow$ *gap between the femur and tibia point of edgeInfo*
    if $JSW < minGap$ **then**
        $jsnBegin \Leftarrow edgeInfo$
    end if
    if $JSW > minGap$ (After the beginning of narrowing start) **then**
        $jsnEnd \Leftarrow edgeInfo$
        if $jsnBegin$ and $jsnEnd$ have values **then**
            $jsnPercentage \Leftarrow$ Calculate based on equation 2
            $jsnData \Leftarrow jsnBegin, jsnEnd, jsnPercentage$
            return $jsnData$
        end if
    end if
end for
return $jsnData$

where JSW_{min} represents the benchmark minimum joint space of a healthy joint, and JSW_{avg} denotes the benchmark average joint space of a healthy joint.

3.5.3 Osteophyte detection in intercondylar region

Osteophyte formation serves as an important radiographic marker for assessing KOA severity. While these bony projections can develop in multiple anatomical locations, current research has predominantly focused on marginal osteophytes. The proposed approach makes a unique contribution by specifically detecting the intercondylar osteophytes, which serve as an important early indicator of KOA development but have received limited attention in previous studies. The intercondylar region presents a complex structure with

Algorithm 6 Osteophyte detection in inter-condylar region.

Require: $xRayEdgeData$ of Regions r4 and r5
Ensure: $OsteophyteData$
1: $slope_{min} \Leftarrow 1$
2: $hieght_{min} \Leftarrow$ benchmark gap of JSN/3
3: $OsteophyteData$=[]
4: **for** $edgeInfo$ in filtered $xRayEdgeData$ **do**
5: Compare the $edgeInfo$ with the previous point and identify the change of direction: upward/downward/unchanged
6: **if** There is a sharp change in direction (from upward to downward) **then**
7: Mark that point as the peak
8: Start from peak
9: **repeat**
10: Access the next edge point backward.
11: Calculate the slope between the **peak** and each edge point (Fig. 8)
12: **until** $slope > slope_{cutoff}$ on both side
13: $h_{left} \Leftarrow$ height between the **peak** and the point identified on the left side.
14: $h_{right} \Leftarrow$ height between the **peak** and the point identified on the right side.
15: $hight = \min(h_{left}, h_{right})$
16: **if** $slope > slope_{cutoff}$ on both side and $height > hieght_{min}$ **then**
17: Mark this growth as osteophyte and store info(peak, the two point at a distance of h from peak)
18: **end if**
19: **end if**
20: **if** The peak is not recognizable - peak touches the femur and merged region is not more than a few pixels **then**
21: Mark this merged section as osteophyte and store data.
22: **end if**
23: **end for**
24: **return** $OsteophyteData$

different curved edges, including normal curves, and curves indicative of osteophytes. We have developed an algorithm to identify osteophytes by analyzing the position in the respective region, slope, and width of these curves. The algorithm for detecting osteophytes is outlined in Algorithm 6. A point on the edge is identified as a peak (line 7) if there is a sharp change in direction in the edge contour. To verify whether this peak corresponds to an osteophyte, we evaluate the slope on both sides of the peak (lines 10–12). A local peak is considered osteophytic only if the slope on both sides exceeds the $slope_{cutoff}$ threshold, indicating a sharp angular deviation. This requirement is formalized in line 16. Figure 8 shows the graphical representation of Alogrithm 6.

The values of $slope_{cutoff}$ and $height_{min}$ are determined based on empirical analysis. While this region may naturally exhibit minor undulations, the general anatomical structure maintains a smooth, gradual profile. Even in cases with irregularities not attributed to osteophytes, the changes in slope tend to be mild. By analyzing a broad set of samples, we observed that osteophytes consistently formed sharper peaks in the edge profile. Based on this insight, we determined a $slope_{cutoff}$ value of 1 to be effective in distinguishing these sharper, pathological peaks from normal variations.

Additionally, if both femur and tibia edges come into contact with each other, where the slope is not measurable, these instances can be identified as well and marked as osteophytes.

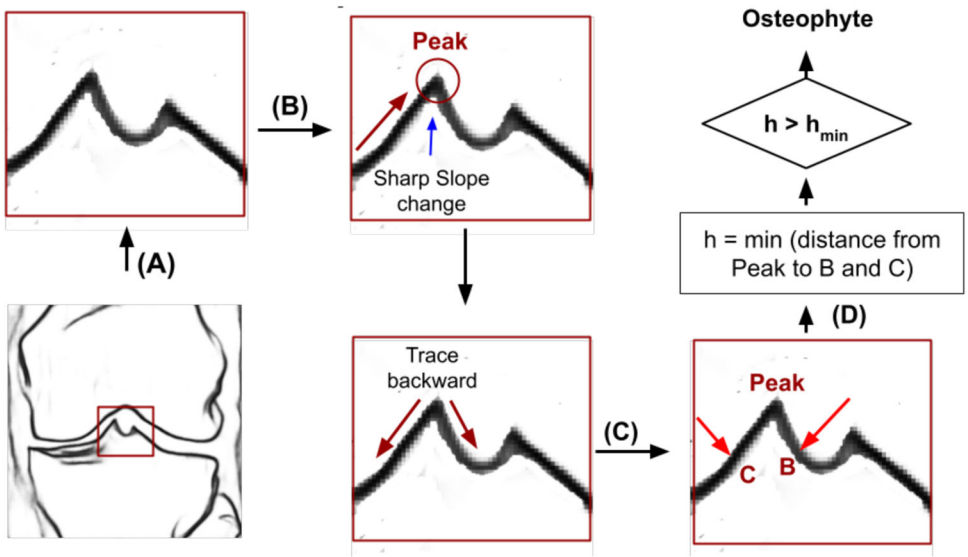


Fig. 8 Identification of osteophytes in the intercondylar region: **A** The edge information of Regions 4 and 5 from xRayEdgeData has been extracted, **B** tracing the edge from the starting of Regions 4 or 5 inward and identifying the point where a sharp change in slope occurs, marking it as the *peak*, **C** tracing backward from the peak to find the point where the slope relative to the peak drops below $slope_{cutoff}$, **D** calculating the vertical distance between the peak and these identified points and selecting the minimum value as the $height$. If $height > height_{min}$, the growth is classified as an osteophyte

Table 1 Performance of edge detection techniques

Edge detection methods	Performance		
	$p = 0.1$, $q = 0.05$	$p = 0.12$, q = 0.04	$p = 0.13$, $q = 0.02$
Canny	0.753125	0.791	0.875375
DexiNed	0.91375	0.8995	0.894625
CLAHE+DexiNed	0.891875	0.9055	0.93875

3.6 Testbed

All experiments are conducted on the Google Colab platform, utilizing a system configuration comprising 107.72 GB of storage and 12.67 GB of RAM. Python 3 is selected as the programming environment due to its simplicity, versatility, and extensive support for relevant scientific and machine learning libraries.

3.7 Performance matrix

For performance evaluation, the metrics used include *precision, recall, F1 score*, and *accuracy*, all of which are derived from the fundamental parameters: *true positive (TP), true negative (TN), false positive (FP)*, and *false negative (FN)*. The corresponding equations for these metrics are provided below. Additionally, the performance of the edge detection algorithm is assessed using a custom-defined metric, presented in Eq, 1.

$$precision = \frac{TP}{TP+FP} \tag{3}$$

$$recall = \frac{TP}{TP+FN} \tag{4}$$

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{5}$$

$$accuracy = \frac{TP+TN}{TP+FN+TN+FP} \tag{6}$$

4 Results and discussion

For testing, we have utilized 450 knee X-ray images from "Knee Osteoarthritis Severity Grading Dataset" [49] from OAI that was annotated and reviewed by an orthopedic surgeon.

4.1 Edge detection

Edge detection of the knee joint from radiographic images constitutes the critical first step in our analysis pipeline, as it fundamentally influences the accuracy of all subsequent processing stages. We have conducted a comprehensive evaluation of various edge detection methodologies, encompassing both traditional image processing techniques and machine learning approaches. A representative subset of the evaluated models is presented in Table 1. Performance assessment has been conducted using Eq. 1, with results averaged across all samples using different weight combinations of parameters p and q. Based on this systematic evaluation, the combination of DexiNed with CLAHE preprocessing has

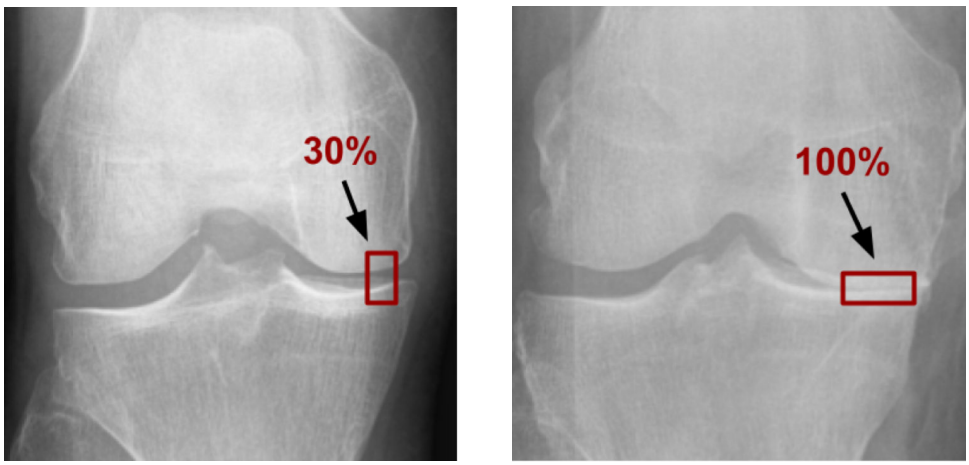
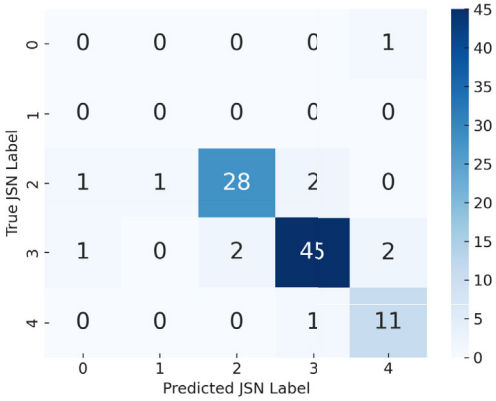


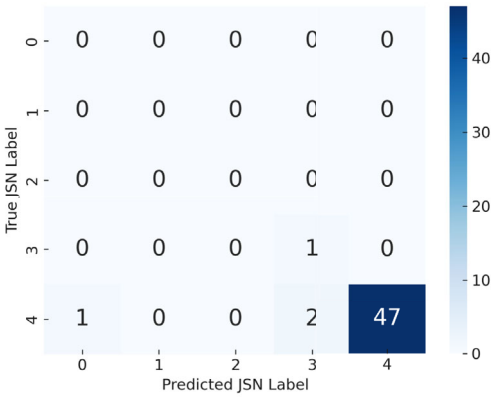
Fig. 9 Detection of JSN: 1^{st} X-ray in **a** has 30% and 2nd X-ray in **b** has 100% JSN

(a) Grade 1

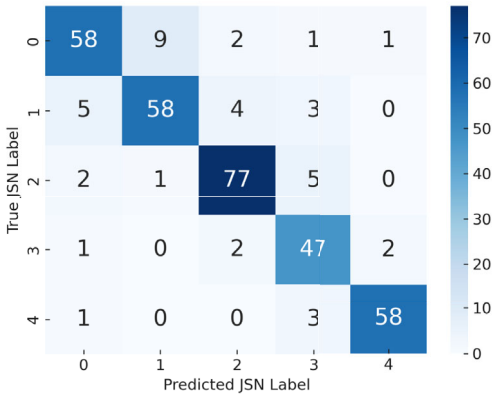
(b) Grade 2



(c) Grade 3



(d) Grade 4



(e) Overall

Fig. 10 Confusion matrices of JSN class at different grade levels (1 to 4) and overall. Each figure shows that each grade is predominantly associated with a specific JSN category, while also showing notable overlap with adjacent categories

emerged as the optimal methodology, demonstrating superior edge detection performance.

As a part of the analysis of the result, we begin with a setting $p = 2q$ and consider the worst-case scenario where all six intended edges are missing and there are up to 10 noisy edges. Applying Eq. 1 under the worst-case scenario assumption (*performance* = 0), the parameters were computed as $p = 0.1$ and $q = 0.05$. However, a missing edge has more than double the impact of a noisy edge. Hence, we have tested different ratios, including $p = 3q$, $p = 4q$, and others. We have calculated the mean performance across all samples and found that when $p > 2q$, DexiNed with CLAHE outperforms all other techniques.

While CLAHE reduces the occurrence of missing edges, it also increases noise. When $p = 2q$, DexiNed performs better than DexiNed with CLAHE due to the higher weight on noise reduction. However, as the weight on missing edges increases (with $p > 2q$), DexiNed with CLAHE shows superior performance by effectively balancing the trade-off between missing edges and noise. Moreover, the noise introduced by CLAHE was alleviated in the subsequent processing stage through a thresholding step.

4.2 Assessment of joint space narrowing

In this study, the JSN has been quantitatively assessed and visually marked on the X-ray images. Figure 9 shows some sample JSN outputs from the proposed JSN detection module. The first X-ray (Fig. 9a) exhibits 30% JSN, and the second X-ray (Fig. 9b) exhibits 100% JSN, complete overlapping of the femur and tibia, indicating advanced stage of KOA.

Figure 10 shows the confusion matrices of each KL grade for different JSN levels. The confusion matrices indicate that each KOA grade is predominantly associated with a specific JSN category, while also showing notable overlap with adjacent categories. For example, JSN category 2 (>25% to 50%) is most common in Grade 2, with a substantial presence of JSN categories 1 and 3. Likewise, JSN category 3 (>50% to 75%) occurs most frequently in Grade 3, along with a noticeable number of cases in JSN categories 2 and 4. Grade 4 exhibits a distinct pattern, being primarily characterized by a single dominant JSN level. In contrast, Grade 1 displays a mix of JSN categories 0 and 1, suggesting the need for more refined or specialized feature extraction techniques to accurately capture its characteristics.

Tables 2 and 3 summarize the JSN performance metrics, categorized by grade and by JSN level, respectively. The JSN-level performance analysis indicates that the boundary categories (JSN 0 and JSN 4) are recognized with the highest precision and recall, reflecting more distinct clinical features. Conversely, the intermediate categories (JSN 2 and JSN 3) demonstrate relatively lower precision and recall, suggesting increased overlap and a higher likelihood of misclassification between moderate stages of JSN.

Table 2 Performance metrics of JSN by each KL grade

Grade	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%) (95% CI)
Grade 1	76	73	74	81.13 (73.74–88.51)
Grade 2	76	76	75	88.89 (82.62–95.16)
Grade 3	75	72	73	87.18 (78.91–95.46)
Grade 4	85	85	85	95.00 (85.20–100.0)
Overall	78	76	77	87.65 (84.15–91.14)

The grade-wise evaluation of the JSN classification model demonstrates consistent performance across Grades 1 to 4, as well as in overall metrics. The assessment includes precision, recall, F1 score, and accuracy with 95% confidence inervals. Grade 4 exhibits the highest performance, achieving 85% in precision, recall, and F1 score, along with 95% accuracy. In contrast, Grade 1 poses greater classification challenges, potentially due to the subtlety of early-stage clinical features, which often result in borderline cases. Despite these challenges, the model maintains strong overall performance, highlighting its robustness in differentiating across varying JSN and severity levels.

Additionally, the minimum, maximum, mean, and standard deviation of JSN values are calculated for the samples within each grade, revealing clear distinctions among them. The average JSN across grades is presented in Fig. 11. These statistics serve as a baseline and are subsequently utilized in formulating the quantitative severity classification approach.

4.3 Osteophyte detection

Figure12 shows a varying number of osteophyte detections, marked with one or more blue rectangle(s). As evident, the system can accurately identify the osteophytes in the inter-condylar region—the most challenging locations for osteophyte detection. Figure 13 summarizes the performance of intercondylar ostephyte detection across different grades. The bar chart clearly demonstrates that the model performs

Table 3 Performance metrics of JSN for each JSN class

JSN Class	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
JSN 0	95	85	89	85
JSN 1	85	80	82	80
JSN 2	70	65	67	65
JSN 3	75	70	72	70
JSN 4	90	85	87	85
Overall	83	77	80	87.65

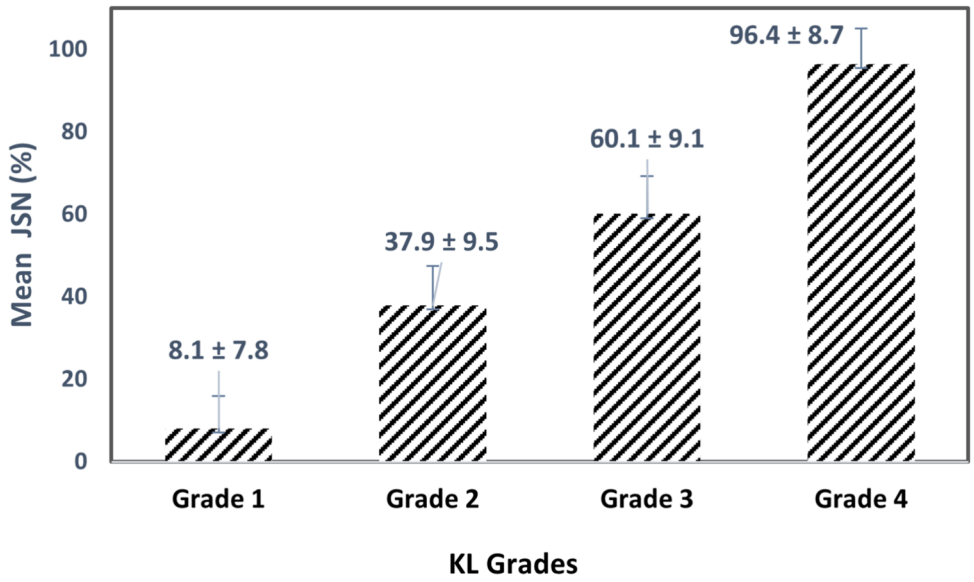


Fig. 11 Bar plot showing the percentage of JSN (mean ± standard deviation) for each grade. Each grade shows a distinct JSN range, lowest in Grade 1 (mild) and highest in Grade 4 (severe)

exceptionally well in detecting Grade 1 osteophytes, achieving a precision, recall, and F1 score of 93.65%, along with an accuracy of 91.58% within a narrow 95% confidence interval. However, performance declines for Grades 2 and 3. Specifically, recall drops to 78.02% for Grade 2, while Grade 3 records the lowest recall (75.95%) and accuracy (70.53%). These findings suggest that mild osteophytes (Grade 1) are distinctly identifiable, whereas the model struggles to differentiate between moderate grades, reflecting the clinical reality where advancing JSN can obscure intercondylar osteophytes. The performance trend observed across KL grades mirrors the disease's natural pathophysiological progression. Nevertheless, these challenges are partly attributed to the inherent limitations of X-ray imaging, especially the possibility of anatomical overlap in two-dimensional views. The accuracy of detection could be improved by incorpo-

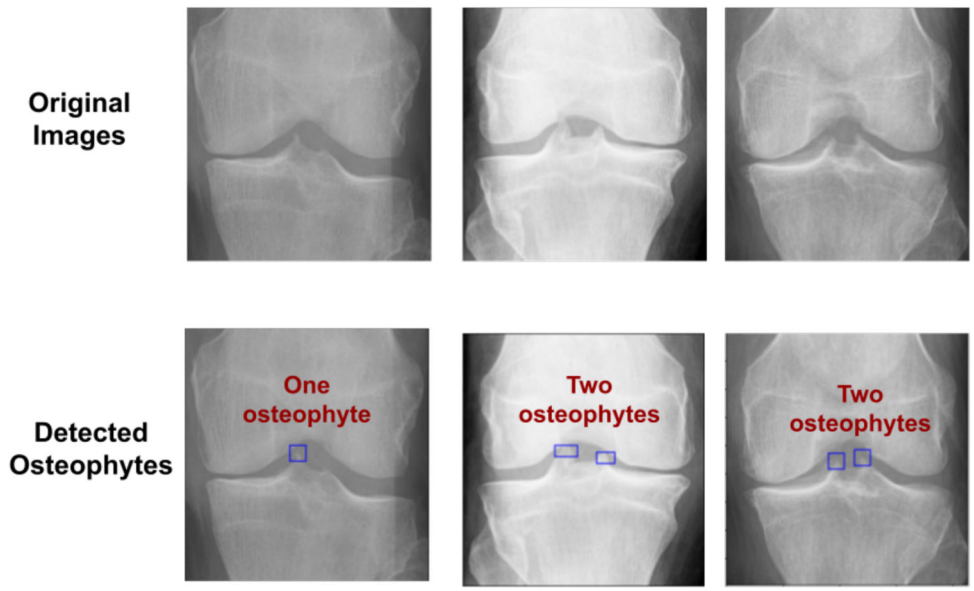


Fig. 12 Detected osteophytes: left X-ray has one osteophyte, middle and right X-ray images have two osteophytes each

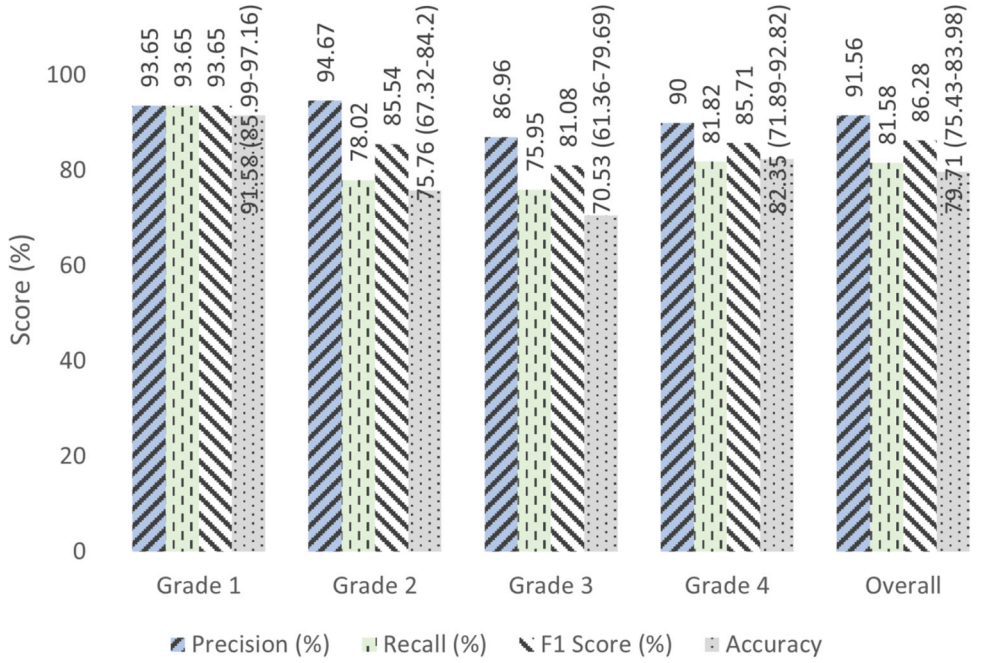


Fig. 13 Performance of osteophyte detection by KL grade and overall

rating additional imaging perspectives, such as lateral and oblique views.

4.4 Severity classification

After analyzing the JSN and intercondylar osteophytes we have observed distinct characteristics for the X-ray images under each grade, as depicted in Table 4. Except for a few outlier data points, the majority of the data conform to these criteria.

Following the classification criteria outlined in Table 4, the proposed system categorizes the severity of the sample images. The model's performance is assessed by comparing its predicted severity levels with the reference grades available in the dataset. The classification result, using precision, recall, F1 score, and accuracy with confidence intervals, across different grades, is illustrated in Table 5. Overall, the model demonstrates balanced performance, with an average precision of 74.56%, recall of 73.76%, F1 score of 73.98%, and accuracy of 72.27% (95% CI: 67.28–76.77%). However, classification accuracy for Grade 2 is comparatively low, which can be attributed to the current system's binary approach to osteophyte detection. While the KL grading system differentiates between "possible osteophyte lipping" (Grade 1) and "definite osteophytes" (Grade 2), our current framework only detects the presence or absence of osteophytes without size differentiation. Future iterations of this research will incorporate osteophyte size estimation, which is expected to enhance classification accuracy and provide more detailed information for clinical decision-making.

Table 4 KL grading based on JSN and intercondylar osteophytes

Grade	Criteria
Grade 1	< 25% JSN with 0 to 1 osteophyte.
Grade 2	25%–50% JSN with 1 or more osteophytes.
Grade 3	51%–75% JSN with 2 or more osteophytes.
Grade 4	> 75% JSN with 2 or more osteophytes or 100% JSN.

Table 5 Performance metrics of KL grade classification with 95% confidence intervals

Grade	Precision (%)	Recall (%)	F1 score (%)
Grade 1	69.23	76.60	72.73
Grade 2	61.54	64.65	63.05
Grade 3	81.48	69.47	75.00
Grade 4	86.00	84.31	85.15
Overall metrics			
Accuracy		72.27%	
95% confidence interval		(67.28%, 76.77%)	
Macro avg. precision		74.56	
Macro avg. recall		73.76	
Macro avg. F1 score		73.98	

4.5 Comparison with existing approaches

Table 6 provides a comprehensive comparison among the existing methodologies and the proposed approach. The majority of available studies in this domain are predominantly based on DL architectures: [43] employs DenseNet and EfficientNetV2; [25] utilizes EfficientNetV2 coupled with an adaptive thresholding strategy; [34] adopts Inception-ResNet-v2 for classification; [42] applies an ensemble approach combining ResNet-101, ShuffleNet-V2, and EfficientNet-V2-s; [47] leverages a hierarchical classification framework, extracting JSN and osteophyte-related geometric features using U-Net, and radiometric features using traditional image processing techniques, subsequently classifying using machine learning models such as logistic regression, support vector machines (SVM), and random forests (RF). [48] (MediaAI-OA) utilizes HRNet for JSN quantification and NASNet for osteophyte detection and KL grade classification; [57] adopted EfficientNet-B5 and DenseNet121 architectures; [58] used a range of pretrained models including VGG16, VGG19, ResNet101, MobileNetV2, Inception-ResNetV2, and DenseNet121; [59] employed VGG-16 and ResNet-50 to distinguish between normal and osteoarthritic knees. Notably, only item 1 primarily relies on purely image processing techniques rather than deep learning.

Our analysis reveals that, aside from the study presented in [48], previous research efforts have not performed quantitative assessment of JSN or osteophytes. While DL-based models offer the advantage of learning complex spatial features (e.g., edges, textures, bone morphology) directly from raw images without manual feature engineering, they exhibit several limitations that impact their clinical applicability. Firstly, they depend heavily on large, balanced, and accurately annotated datasets, and their performance can decline in cases of data scarcity or class imbalance. Secondly, both training and inference are resource-intensive, requiring powerful GPUs and significant memory capacity. Thirdly, DL models often lack interpretability, which poses challenges for clinical validation and can limit trust among medical professionals. Although some studies integrate explainable AI (XAI) tools to highlight important regions/features contributing to decisions, these works often lack the quantitative evaluation of the highlighted areas.

Our proposed model differs from the conventional DL-based pipelines. It avoids extensive training by utilizing a pretrained model in the zero-shot inference setting to perform the X-ray edge extraction. Subsequent stages of our model employ inexpensive image processing techniques and a pixel-wise scanner to detect osteophyte and JSN features, making the system lightweight and deployable on standard CPUs. In addition, our proposed model does not rely on large annotated datasets, making it suitable for any low-resource setting. The only annotation that is needed for measuring the performance of the model.

The proposed system demonstrates robust accuracy in JSN quantification. Although osteophyte detection accuracy is moderately lower compared to JSN performance, this system is specifically designed to detect early osteophytes in

Table 6 Comparison of the proposed model with other existing methods

References	Accuracy			Comments
	JSN	Osteophytes	Classification	
Saleem et al. [22]	X	X	97.14	Classified based on JSN only and binary classification
Phan et al. [43]	X	X	89.34	Binary classification with highly imbalanced dataset
Prezja et al. [25]	X	X	65.48	Needs labeled data and no quantitative result.
M and Goswami [34]	X	X	91.03	Same as above
Pi et al. [42]	X	X	76.33	Same as above
Pan et al. [47]	X	X	65.98	Same as above
Mohammed et al. [58]	X	X	89	Same as above
Alshamrani et al. [59]	X	X	92	Same as above
Ahmed and Imran [60]	X	X	67	Same as above
Yoon et al. [48]	X	84	83	Used different supervised model. Measured the JSN and marginal osteophytes but didn't use those for classification.
Proposed Method	88	80	73	Osteophytes in the intercondylar area were detected. Classification is based on the quantitative data of JSN and osteophytes.

the intercondylar region, an area often overlooked in existing automated KOA detection frameworks. Although medical studies have discussed the clinical significance of intercondylar osteophytes, to the best of our knowledge, no automated system has been previously developed to specifically identify these osteophytes. This novel contribution advances the field of automated KOA assessment. While the accuracy of the presented system is modest relative to some existing approaches, its primary contribution lies in three key areas: providing quantitative measurements of radiographic features, handling error cases effectively, and eliminating the expensive training and its dependency on labeled datasets.

4.6 Limitations

This research provides significant contributions to automated KOA diagnosis, while acknowledging certain limitations that present opportunities for future development. The current framework focuses specifically on intercondylar osteophytes, without incorporating marginal osteophytes in the assessment process. This limitation potentially affects the overall accuracy of severity grading. Our ongoing research efforts are focused on developing a more robust method for detecting and quantifying both intercondyler and marginal osteophytes to enhance the classification accuracy.

While our framework provides robust analysis of radiological features, we acknowledge the importance of clinical symptoms in comprehensive KOA diagnosis. The National Institute of Health and Care Excellence (NICE) guidelines [16] emphasize that diagnosis should consider multiple factors, particularly in individuals aged 45 or above who experience activity-related pain with specific patterns of morning stiffness. Future research directions include the integration of clinical symptoms into our assessment framework. This expansion would create a more comprehensive diagnostic approach by combining both imaging biomarkers and clinical indicators, potentially leading to improved diagnostic precision.

The performance of the proposed model is strongly dependent on the accurate detection of tibial and femoral surface edges. However, X-ray imaging inherently poses challenges such as edge blending with adjacent soft tissues, leading to potential incomplete or noisy edge extraction. Additionally, the two-dimensional nature of X-ray imaging can cause anatomical overlaps, which may obscure important structural features. To overcome these limitations, future research will focus on integrating additional images, including lateral and oblique views, and employing ensemble strategies along with advanced edge enhancement techniques to further improve the detection accuracy.

For ROI localization, the template matching algorithm is employed due to its simplicity and computational efficiency, although it has inherent limitations in handling anatomical variability. Specifically, the approach is sensitive to differences in knee size, orientation, and positioning across subjects, which may affect localization accuracy. To address this, future work will focus on incorporating more robust and adaptive ROI detection methods using DL object detection models such as Faster R-CNN, SSDs, and YOLO.

5 Conclusion

Knee osteoarthritis, a prevalent and progressively degenerative disease, can substantially impair mobility and may lead to disability if not addressed early. This study introduces a novel framework for early-stage KOA detection by quantitatively assessing two key diagnostic radiographic markers: joint space narrowing and intercondylar osteophytes. Most notably, we present the first automated system to quantify osteophytes in the intercondylar region and correlate these measurements with automatic KOA severity grading. The proposed system demonstrates significant accuracy in early-stage disease detection while minimizing inter- and intra-observer variability. Furthermore, the resulting quantitative data can be leveraged within alternative OA grading systems, offering a significant advancement over traditional qualitative methods.

Author Contributions TH: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, writing—original draft, writing—review and editing. TF: conceptualization, investigation, methodology, project administration, supervision, validation, writing—review and editing. AB: conceptualization, data curation, investigation, methodology, project administration, supervision, validation, writing—review and editing. MN: conceptualization, investigation, methodology, project administration, supervision, validation, writing—review and editing.

Declarations

Competing interests The authors declare no competing interests.

References

1. Long H, Liu Q, Yin H, Wang K, Diao N, Zhang Y, Lin J, Guo A (2022) Prevalence trends of site-specific osteoarthritis from 1990 to 2019: findings from the global burden of disease study 2019. Arthritis Rheumatol 74(7):1172–1183. https://doi.org/10.1002/art.42089. Epub 2022 Jun 2
2. Andrianakos AA, Kontelis LK, Karamitsos DG, Aslanidis SI, Georgountzos AI, Kaziolas GO, Pantelidou KV, Vafiadou EV, Dantis PC, Group ES (2006) Prevalence of symptomatic knee, hand, and hip osteoarthritis in Greece. The ESORDIG study. J Rheumatol 33(12):2507–2513
3. SingleCare Team (2024) Arthritis statistics 2024: what percent of the population has arthritis? https://www.singlecare.com/blog/news/arthritis-statistics/. Accessed 9 Aug 2024

4. Steinmetz JD, Culbreth GT, Haile LM, Rafferty Q, Lo J, Fukutaki KG, Cruz JA, Smith AE, Vollset SE, Brooks PM et al (2023) Global, regional, and national burden of osteoarthritis, 1990–2020 and projections to 2050: a systematic analysis for the global burden of disease study 2021. The Lancet Rheumatol 5(9):508–522
5. Berrimi M, Hans D, Jennane R (2024) A semi-supervised multiview-MRI network for the detection of Knee Osteoarthritis. Comput Med Imaging Graph 114:102371. https://doi.org/10.1016/j.compmedimag.2024.102371
6. Huo J, Ouyang X, Si L, Xuan K, Wang S, Yao W, Liu Y, Xu J, Qian D, Xue Z, Wang Q, Shen D, Zhang L (2022) Automatic grading assessments for knee MRI cartilage defects via self-ensembling semi-supervised learning with dual-consistency. Med Image Anal 80:102508. https://doi.org/10.1016/j.media.2022.102508
7. Ambellan F, Tack A, Ehlke M, Zachow S (2019) Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the osteoarthritis initiative. Med Image Anal 52:109–118. https://doi.org/10.1016/j.media.2018.11.009
8. Segal NA, Akkaya Z, Jeon JH, Turmezei T (2024) Semi-quantitative weight-bearing assessment of knee osteoarthritis: COAKS (CT Osteoarthritis Knee Score) reliability. Osteoarthr Imaging 4(3):100243. https://doi.org/10.1016/j.ostima.2024.100243
9. Burlison JM, Bowes MA, Conaghan PG, Brett AD (2024) 3D bone shape from CT-scans provides an objective measure of osteoarthritis severity: data from the IMI-APPROACH study. Osteoarthr Imaging 4(4):100250. https://doi.org/10.1016/j.ostima.2024.100250
10. Emadi Andani M, Salehi Z (2024) An affordable and easy-to-use tool to diagnose knee arthritis using knee sound. Biomed Sig Process Control 88:105685. https://doi.org/10.1016/j.bspc.2023.105685
11. Khan TI, Sakib N, Hassan MM, Ide S (2024) Gaussian mixture model in clustering acoustic emission signals for characterizing osteoarthritic knees. Biomed Sig Process Control 87:105510. https://doi.org/10.1016/j.bspc.2023.105510
12. Kellgren JH, Lawrence J (1957) Radiological assessment of osteo-arthrosis. Ann Rheum Dis 16(4):494
13. Ahlback S (1968) Osteoarthrosis of the knee. A radiographic investigation. Acta Radiol Diagn (Suppl) 277:7–72
14. Altman RD, Hochberg M, Murphy W Jr, Wolfe F, Lequesne M (1995) Atlas of individual radiographic features in osteoarthritis. Osteoarthr Cartil 3:3–70
15. Hefti F, Müller W, Jakob R, Stäubli H-U (1993) Evaluation of knee ligament injuries with the IKDC form. Knee Surg Sports Traumatol Arthroscopy 1(3–4):226–234
16. NICE (2023) Osteoarthritis in over 16s: diagnosis and management - Recommendations. https://www.nice.org.uk/guidance/ng226/chapter/Recommendations#diagnosis. Accessed: 27 Aug 2024
17. Sasho T, Akagi R, Tahara M, Katsuragi J, Nakagawa R, Enomoto T, Yusuke S, Kimura S, Yamaguchi S, Watanabe A (2017) Osteophyte formation on medial wall of the intercondylar notch of femur is an early sign of osteoarthritic knee development using osteoarthritis initiative data. Osteoarthr Cartil 25:33. https://doi.org/10.1016/j.joca.2017.02.066
18. Sasho T, Akagi R, Tahara M, Watanabe A, Ogawa Y, Kimura S, Sadamasu A, Yamaguchi S (2018) Frequency of osteophyte formation at posterior notch of the femur among healthy knees of 40's and 50's. Osteoarthr Cartil 26:198. https://doi.org/10.1016/j.joca.2018.02.421
19. Patron A, Annala L, Lainiala O, Paloneva J, Äyrämö S (2022) An automatic method for assessing spiking of tibial tubercles associated with knee osteoarthritis. Diagnostics 12(11). https://doi.org/10.3390/diagnostics12112603
20. Shamir L, Ling SM, Scott WW Jr, Bos A, Orlov N, Macura TJ, Eckley DM, Ferrucci L, Goldberg IG (2009) Knee X-ray image analysis method for automated detection of osteoarthritis. IEEE Trans Biomed Engi 56(2):407–415. https://doi.org/10.1109/TBME.2008.2006025
21. Wright RW, Group M (2014) Osteoarthritis classification scales: interobserver reliability and arthroscopic correlation. J Bone Joint Surg Am Vol 96(14):1145–1151. https://doi.org/10.2106/JBJS.M.00929
22. Saleem M, Farid MS, Saleem S, Khan MH (2020) X-ray image analysis for automated knee osteoarthritis detection. Signal Image Video Process 14:1079–1087
23. Wang Y, Wang X, Gao T, Du L, Liu W (2021) An automatic knee osteoarthritis diagnosis method based on deep learning: data from the osteoarthritis initiative. J Healthc Eng 2021:1–10
24. Mahum R, Rehman SU, Meraj T, Rauf HT, Irtaza A, El-Sherbeeny AM, El-Meligy MA (2021) A novel hybrid approach based on deep CNN features to detect knee osteoarthritis. Sensors 21(18):6189
25. Prezja F, Annala L, Kiiskinen S, Lahtinen S, Ojala T (2023) Adaptive variance thresholding: a novel approach to improve existing deep transfer vision models and advance automatic knee-joint osteoarthritis classification. arXiv:2311.05799
26. Chen P, Gao L, Shi X, Allen K, Yang L (2019) Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss. Comput Med Imaging Graph 75:84–92
27. Sadeghzadehyazdi N, Batabyal T, Acton ST (2021) Modeling spatiotemporal patterns of gait anomaly with a CNN-LSTM deep neural network. Expert Syst Appl 185:115582. https://doi.org/10.1016/j.eswa.2021.115582
28. Dupuis DE, Beynnon BD, Richard MJ, Novotny JE, Skelly JM, Cooper SM (2003) Precision and accuracy of joint space width measurements of the medial compartment of the knee using standardized MTP semi-flexed radiographs. Osteoarthr Cartil 11(10):716–724. https://doi.org/10.1016/S1063-4584(03)00158-4
29. Mazzuca SA, Brandt KD, Buckwalter JA, Lequesne M (2004) Pitfalls in the accurate measurement of joint space narrowing in semi-flexed, anteroposterior radiographic imaging of the knee. Arthritis Rheum 50(8):2508–2515. https://doi.org/10.1002/art.20363
30. Paixao T, DiFranco MD, Ljuhar R, Ljuhar D, Goetz C, Bertalan Z, Dimai HP, Nehrer S (2020) A novel quantitative metric for joint space width: data from the Osteoarthritis Initiative (OAI). Osteoarthr Cartil 28(8):1055–1061. https://doi.org/10.1016/j.joca.2020.04.003
31. Marijnissen AC, Vincken KL, Vos PA, Saris DB, Viergever MA, Bijlsma JW, Bartels LW, Lafeber FP (2008) Knee Images Digital Analysis (KIDA): a novel method to quantify individual radiographic features of knee osteoarthritis in detail. Osteoarthr Cartil 16(2):234–243. https://doi.org/10.1016/j.joca.2007.06.009
32. Oka H, Muraki S, Akune T, Mabuchi A, Suzuki T, Yoshida H, Yamamoto S, Nakamura K, Yoshimura N, Kawaguchi H (2008) Fully automatic quantification of knee osteoarthritis severity on plain radiographs. Osteoarthr Cartil 16(11):1300–1306. https://doi.org/10.1016/j.joca.2008.03.010
33. Poma XS, Riba E, Sappa A (2020) Dense extreme inception network: towards a robust CNN model for edge detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp 1923–1932
34. M, GK, Goswami AD (2023) Automatic classification of the severity of knee osteoarthritis using enhanced image sharpening and CNN. Appl Sci 13(3). https://doi.org/10.3390/app13031658
35. Sikkandar MY, Begum SS, Alkathiry AA, Alotaibi MSN, Manzar MD (2022) Automatic detection and classification of human knee osteoarthritis using convolutional neural networks. Comput Mater Continua 70(3):4279–4291

36. Yunus U, Amin J, Sharif M, Yasmin M, Kadry S, Krishnamoorthy S (2022) Recognition of knee osteoarthritis (KOA) using YOLOv2 and classification based on convolutional neural network. Life 12(8):1126
37. Swiecicki A, Li N, O'Donnell J, Said N, Yang J, Mather RC, Jiranek WA, Mazurowski MA (2021) Deep learning-based algorithm for assessment of knee osteoarthritis severity in radiographs matches performance of radiologists. Comput Biol Med 133:104334. https://doi.org/10.1016/j.compbiomed.2021.104334
38. El-Ghany SA, Elmogy M, El-Aziz AAA (2023) A fully automatic fine tuned deep learning model for knee osteoarthritis detection and progression analysis. Egypt Inform J 24(2):229–240. https://doi.org/10.1016/j.eij.2023.03.005
39. Alshamrani HA, Rashid M, Alshamrani SS, Alshehri AHD (2023) Osteo-net: an automated system for predicting knee osteoarthritis from x-ray images using transfer-learning-based neural networks approach. Healthcare (Basel) 11(9):1206. https://doi.org/10.3390/healthcare11091206
40. Touahema S, Zaimi I, Zrira N, Ngote MN, Doulhousne H, Aouial M (2024) MedKnee: a new deep learning-based software for automated prediction of radiographic knee osteoarthritis. Diagnostics 14(10). https://doi.org/10.3390/diagnostics14100993
41. Tariq T, Suhail Z, Nawaz Z (2023) Knee osteoarthritis detection and classification using X-rays. IEEE Access 11:48292–48303. https://doi.org/10.1109/ACCESS.2023.3276810
42. Pi S-W, Lee B-D, Lee M, Lee H (2023) Ensemble deep-learning networks for automated osteoarthritis grading in knee X-ray images. Sci Rep 13. https://doi.org/10.1038/s41598-023-50210-4
43. Phan TH, Nguyen TT, Nguyen TD, et al (2025) DIKOApp: an AI-based diagnostic system for knee osteoarthritis. J Digit Imaging Inform Med. https://doi.org/10.1007/s10278-024-01383-5
44. Fatema K, Rony MA, Azam S, Hossain MS, Karim A, Hasan M, Jonkman M (2023) Development of an automated optimal distance feature-based decision system for diagnosing knee osteoarthritis using segmented x-ray images. Heliyon 9:21703. https://doi.org/10.1016/j.heliyon.2023.e21703
45. Ribas LC, Riad R, Jennane R, Bruno OM (2022) A complex network based approach for knee osteoarthritis detection: data from the osteoarthritis initiative. Biomed Sig Process Control 71:103133. https://doi.org/10.1016/j.bspc.2021.103133
46. Ahmed SM, Mstafa RJ (2022) Identifying severity grading of knee osteoarthritis from X-ray images using an efficient mixture of deep learning and machine learning models. Diagnostics 12(12). https://doi.org/10.3390/diagnostics12122939
47. Pan J, Wu Y, Tang Z et al (2024) Automatic knee osteoarthritis severity grading based on X-ray images using a hierarchical classification method. Arthritis Res Ther 26:203. https://doi.org/10.1186/s13075-024-03416-4
48. Yoon JS, Yon C-J, Lee D, Lee JJ, Kang CH, Kang S-B, Lee N-K, Chang CB (2023) Assessment of a novel deep learning-based software developed for automatic feature extraction and grading of radiographic knee osteoarthritis. BMC Musculoskelet Disord 24(1). https://doi.org/10.1186/s12891-023-06951-4
49. Chen P (2018) Knee osteoarthritis severity grading dataset. https://data.mendeley.com/datasets/56rmx5bjcr/1, https://doi.org/10.17632/56rmx5bjcr.1
50. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9(7):671–675. https://doi.org/10.1038/nmeth.2089
51. Pizer SM, Johnston RE, Ericksen JP, Yankaskas BC, Muller KE (1990) Contrast-limited adaptive histogram equalization: speed and effectiveness. In: [1990] Proceedings of the first conference on visualization in biomedical computing. pp 337–345. https://doi.org/10.1109/VBC.1990.109340
52. Xie S, Tu Z (2015) Holistically-nested edge detection. In: Proceedings of the IEEE international conference on computer vision. pp 1395–1403
53. Ziou D, Tabbone S (1998) Edge detection techniques-an overview. Pattern Recognit Image Anal Adv Math Theory Appl 8(4):537–559
54. Dharampal VM (2015) Methods of image edge detection: a review. J Electr Electron Syst 4(2):5
55. Soria X, Pomboza-Junez G, Sappa AD (2022) LDC: lightweight dense CNN for edge detection. IEEE Access 10:68281–68290
56. Itseez (2022) Opencv-python. https://pypi.org/project/opencv-python/. Accessed 15 May 2024
57. Sinha V, Veemaraj E (2024) Knee osteoarthritis detection using X-rays and DNN. In: 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC). pp 467–472. https://doi.org/10.1109/ICAAIC60222.2024.10575881
58. Mohammed A, Hasanaath A, Latif G, Bashar A (2023) Knee osteoarthritis detection and severity classification using residual neural networks on preprocessed X-ray images. Diagnostics (Basel) 13(8):1380. https://doi.org/10.3390/diagnostics13081380
59. Alshamrani HA, Rashid M, Alshamrani SS, Alshehri AHD (2023) Osteo-net: an automated system for predicting knee osteoarthritis from X-ray images using transfer-learning-based neural networks approach. Healthcare 11(9). https://doi.org/10.3390/healthcare11091206
60. Ahmed R, Imran AS (2024) Knee osteoarthritis analysis using deep learning and XAI on X-rays. IEEE Access 12:68870–68879. https://doi.org/10.1109/ACCESS.2024.3400987

Tanjina Helaly is pursuing Ph.D. in Computer Science and Engineering at Bangladesh University of Engineering and Technology (BUET). She is also working as a faculty at the Department of Computer Science and Engineering, University of Asia Pacific (UAP), Bangladesh. She completed her Masters in Computer Science from North Dakota State University, USA and she has a sound industry experiences. Her research focuses on advancing machine learning in healthcare, with a strong emphasis on early orthopedic disease prediction.

Tanvir R. Faisal is working an Assistant Professor at the University of Louisiana at Lafayette, USA. He is the director of Musculoskeletal Mechanics & Multiscale Materials (4M) Laboratory. His research includes image-based biomechanical modeling, machine learning, and predictive modeling for orthopedic disease detection, finite element analysis, etc. He has many publications in reputed journals based on his work on mathematical and computational modeling of knee joints, cartilage and ligaments, machine learning/deep learning based osteoporotic fracture risk assessment, etc.

Ahmed Suparno Bahar Moni is an Orthopedic Surgeon and Assistant Professor of Orthopedics and Medical Education department at the University of Toledo, USA. He graduated from Dhaka Medical College, Bangladesh. He has a working experiences in Orthopedics in different countries like Korea, Hong Kong, and Malaysia. He has published several books on musculoskeletal disorders for practicing surgeons and residents.

Mahmuda Naznin is working as a Professor, and also served as the Head of the Department of Computer Science and Engineering (CSE), Bangladesh University of Engineering and Technology (BUET). She has published more than 50 research articles in peer-reviewed journals published by Springer, IEEE, Elsevier, and in proceedings of international conferences by IEEE, ACM, and Springer. Her research includes applied machine learning, medical imaging, network resource optimization, graph algorithms.