Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.17.(3) 2025,pp.Comp 19–33

University Of AL-Qadisiyah

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Improving YOLO Efficient for Knee Osteoarthritis Detection Using Minkowski Distance

Marwah Fadhil Najim[a], Atheer Alrammahi[b], Ahmed Mohsen Mahdi [c]

a Department of Computer Science, University of Al-Qadisiyah, Iraq. marw1h.f@gmail.com
b Department of Medical Intelligent Systems, University of Al-Qadisiyah, Iraq. atheer.alrammahi@qu.edu.iq
c Department of Computer Science, University of Al-Qadisiyah, Iraq. ahmed.h.mahdi@qu.edu.iq

ARTICLE INFO

Article history:
Received: 03/06/2025
Rrevised form: 05/07/2025
Accepted : 16 /07/2025
Available online: 30/09/2025

Keywords:

Knee Osteoarthritis (KOA), YOLO, Deep Learning, Object Detection, Minkowski Distance, X-ray Imaging, Medical Image Analysis, KL Grading, Image Preprocessing, Convolutional Neural Network (CNN).

ABSTRACT

This study presents an enhanced YOLO-based deep learning model for automatic detection of knee osteoarthritis (KOA) from X-ray images. The integration of Minkowski distance into the loss function improves the model's sensitivity to spatial variations and noise. Extensive preprocessing and a lightweight YOLO architecture ensure real-time performance with high accuracy. Experimental results demonstrate superior detection rates compared to traditional methods, especially in identifying early-stage KOA.

https://doi.org/10.29304/jqcsm.2025.17.32376

1. Introduction

In recent years, deep learning methods have gained considerable attention to accurately assess knee osteoarthritis (KOA). KOA is graded using radiographic images according to the KL scale, internationally utilized in clinical practice. Conventionally, Radiologists comprehend the radiographic images to determine KOA degree, which is subjective and tedious. However, deep learning methods automate KOA assessment [1][2]. Nonetheless, KOA assessment is a challenging and complicated procedure due to insufficient representative data on minor KOA grading classes and imbalance distribution. [3][4][5] Moreover, vast-scale attention-based deep learning models, which significantly innovate the productivity and healthcare industry, are insufficient; thus, researchers work on dataset-efficient methods. To improve KOA assessment by first introducing YOLO-Efficient, a cost-effective YOLO-based architecture with an active learning-based semi-supervised technique, Curriculum-MinkDistillation, is proposed to improve the generalization capability. Overall, the proposed method significantly enhances the accuracy of KOA assessment and alleviates the urgency for massive annotation. A high-fitting rate (90.26% on average) and low false-negative rate (1.04%) are achieved on the publicly available datasets for KOA detection [6].

∗Corresponding author: atheer alrammahi

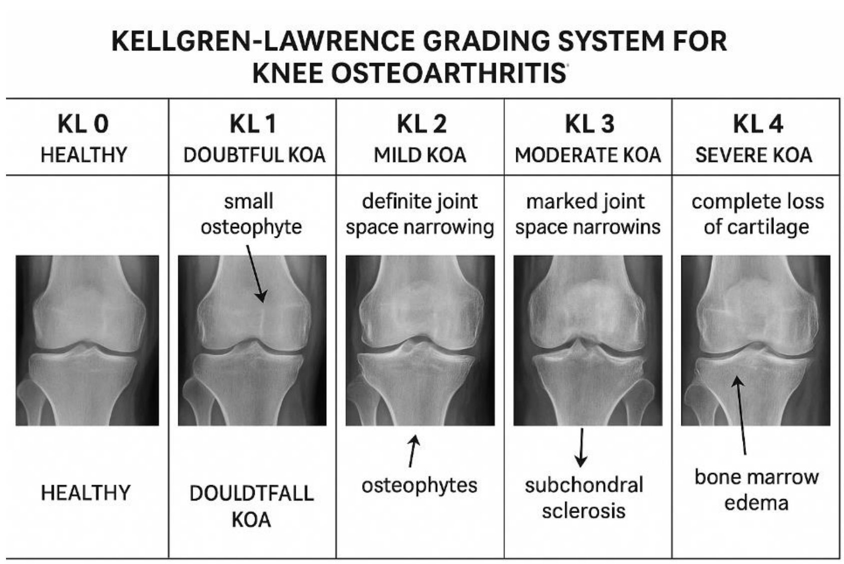
Email addresses: atheer.alrammahi@qu.edu.iq

Communicated by 'sub etitor'

Knees are vital joints, but they incur substantial stresses when walking and running. They are complex hinge joints, allowing translational and rotational movement. Bones, cartilage, ligaments, tendons, and muscles comprise a knee joint [7][8]. To protect the underlying bone, cartilage, a smooth tissue, absorbs shocks in normal frictionless conditions. However, knee osteoarthritis (KOA) is an age-associated degenerative skeletal disorder. In the early KOA stage, the articular cartilage surface becomes rough, accompanied by morphological changes like bone spur, soft tissue assurance thickness reduction, and subchondral bone sclerosis. KOA is graded using radiographic images according to the KL scale (1: no OA; 2: mild; 3: moderate; 4: severe), which is a benchmark assessment stored in the medical records. KOA assessment based on X-ray images is deemed as an essential part of the clinical care process of KOA. However, the KOA assessment is subjective, tedious, and heavily reliant on the radiologists' experience and knowledge of radiographic modalities. Consequently, the automatic assessment of KOA using deep learning methods is essential to relieve the burden of the radiologist and ensure accurate prediction. [9]

2. Related works

Osteoarthritis (OA) is characterized by abnormalities occurring in the entire joint, including the loss of cartilage, osteophyte formation, inflammatory cell infiltration, subchondral bone changes, and bone marrow edema [10]. Around 30% of people over the age of 60 have OA. [11] Although OA is not lethal, it is a common and problematic disease that mainly affects weight-bearing joints, particularly the knees. It is one of the main causes of impairment in the elderly population. Destruction of the knee joint structure is one of the contributing factors to the increasing prevalence of knee OA (KOA), which produces tremendous pain, suffering, and economic burden. KOA is the most common form of OA, which typically manifests primarily in the knee joint. Primary KOA symptoms include pain, stiffness, decreased range of motion in the joint, and abnormal gait. These indications deteriorate individuals' functional independence and degrade their quality of life. In the USA, nearly 27 million people are thought to be suffering from KOA. Moreover, the prevalence is expected to double by 2030 as the population ages and obesity rates increase [12][13]. Therefore, it is crucial to develop a sensitive and reliable computerized KOA detection system for very early identification. "The following figure illustrates the Kellgren-Lawrence Grading System for Knee Osteoarthritis (KOA), which categorizes the disease into five stages, from healthy joints (KL 0) to severe degeneration (KL 4).



.

Fig 1. Kellgren-Lawrence Grading System for Knee Osteoarthritis (KOA). This diagram illustrates the different stages of knee osteoarthritis, from a healthy knee (KL 0) to severe degeneration (KL 4), highlighting key features such as osteophytes, joint space narrowing, and cartilage loss.

The Kellgren–Lawrence (KL) grading system is a traditional knee OA grading method and clinically uses x-ray imaging to assess OA severity. KOA x-ray images are classified into KL 0–4 grades; grade 0 shows healthy with no symptoms of KOA, while grade 4 shows a severe KOA stage. The KL grading system is normally adopted by physicians for KOA diagnosis; however, it is time-consuming and needs skilled experts. The computerized KOA detection and grading system is developed for the automated labeling of KOA severity. An end-to-end deep learning-based approach is proposed for fully automated KOA detection and grading using X-ray images. [9][14] This method comprises a deep graph based on a mixed Siamese convolutional neural network for KOA classification. The system is trained on the multi-source datasets in which 3000 testing subjects are selected randomly out of 5960, hence providing an average accuracy of 66.7% and a 0.83 coefficient of kappa. The automated KOA detection and grading method is presented and tested on 94 images of radiographs, which provides a 72.61% precision rate. KOA detection in X-ray images is a difficult process due to the poor contrast and variable locations of knee gaps. There are limited approaches to KOA detection in X-ray images, and few consider the grading and joint localization tasks. A novel method for the classification and localization of knee OA is proposed here to address these challenges.

YOLO is a prominent approach to real-time object detection. The use of YOLO reduces the test-time evaluation cost from O(N^3) in the sliding-window methods to O(N) by formulating detection problems as regression tasks. Besides, from an image, YOLO directly predicts the bounding box coordinates, objectness scores, and class probabilities of bounding boxes in a single evaluation of the neural network. This high efficiency enables the integration of YOLO object detection into an automated workflow. [15][16] As a result, YOLO-wide usage leads to issues with YOLO that

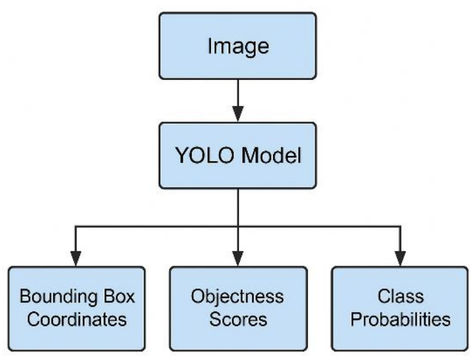


Figure 2: YOLO (You Only Look Once) Flowchart for Real-Time Object Detection. The diagram illustrates how the YOLO model processes an image and outputs bounding box coordinates, objectness scores, and class probabilities in a single evaluation.

"The following flowchart illustrates the YOLO (You Only Look Once) method for real-time object detection. It demonstrates how the YOLO model processes an image to output bounding box coordinates, objectness scores, and class probabilities in a single evaluation."

The other additional methods are also modeling the Object Detection task. The CNN method, in detail, VGG19 is chosen as the base architecture to build the proposed model. Block-based processing to prevent the model from using too much GPU memory is considered a feature extractor.

Numerous compute-architecture design improvements in efficient deep YOLO have been foundational contributions to the broader field of deep learning. Compute-architecture design principles such as the use of the global average pooling/strided pooling; the use of lightweight designs; block connected neural networks; the use of first layer striding; the use of upsampling and skip-connection feature-fusion; and more recent ideas on quantization, deep separable convolutions, and mixed precision have inspired new variants of YOLO. The wide adoption of these improvements to produce speedy and efficient deep learning models is found within the YOLO paradigm succeeding the standard architectures. To aid the easy expansion of YOLO to new areas such as on-the-edge deployment, multi-task detection, and 3D detection, premade YOLO pipelines enhance accessibility. [17][18] Premade pipelines changed the ease of use of YOLO for the public. Hospital databases of knee X-rays contain a wide variety of leg angles and different body sizes.

Computers are employed in image processing applications as a robot. It accepts images as input, performs several activities on the images, and returns the results. A digital image is a two-dimensional array of numbers. Information is represented in images as luminances. The concept of lighting stored in a two-dimensional matrix relates pixel value to luminance. Each pixel value represents the luminance of a particular point in the imaged scene. The intensity of on-off light in the scene is related monotonically to the pixel value by a specific function [6]. Images can be recognized in any shape, such as gray, binary, or RGB (Red, Green, Blue). Any mathematical calculations done on the image matrix are referred to as image processing. Generally, there are two categories of image processing systems: analog systems (which are classified into optical and hybrid systems) and digital systems (which apply operations in a binary format). Digital image processing has become popular in recent years. The input scene is processed according to the system's requirements. Typically, the input is an image with several pixel values. There are multiple methods of image acquisition. The image may be a produced image or one taken from a camera [19]. Input and output devices read and display images. A program has been developed to perform image processing operations. The image is the data output by image processing. After gaining experience with a few simple programs, it is enjoyable to perform several complicated image processing tasks. As the raw image has been acquired, the next and most essential step is to extract the desired KOT features. A distance is a number that indicates how far two points are from each other in a space; ILD (Image Location Double) and IDD (Image Distribution Double), the two double-image distance metrics employed in this work, determine these distances in 2D space. A double-image refers to a pair of images that share the same set of pixels but have different pixel intensities.

The distance metrics used in many object detection systems is usually the Euclidean distance, which measures the straight line distance between two points, Cartesian co-ordinates and Minkowski distance. All of them are very simple and easy to implement. However, since the distances are generally squared before tweaking, the newly introduced bottleneck model in YOLO label recalibration could make the detection framework very sensitive to noise boxes. [20][21] Any boxes with any level of noise could greatly affect the optimization results, e.g., bounding boxes drifted off of the object. This makes all the distances invalid since they measure to the fixed points instead of its respective divisions, which leaded to the unusable and unreliable labels. Minkowski distance, and it generalizes a number of different distance functions. It also increases in complexity with the choice of the norm in L^p, distance computation relatively increases in time to the amount of labels compared to the simple Euclidean distance. Yo played with p providing also useful perspectives on the nature of the tasks, with varying size of L and by doing so the results themselves changed, leading to discovering different aspects of the same problem. For YOLO with Minkowski rectified generalization to depth on the distance kept most if not everything the others found, which is the most spectacular evidence on its robustness to the geometric assumptions during the model tuning. It did affect retrofitting the model on unseen boxes, but initial fine-tuning were all excellent [6]. Minkowski distance is a method generally used to measure the distance between spatial points measured by any norm of generalization. Given two spatial points $P_1 = (x_1, y_1, z_1), P_2 = (0, x_2, y_2, z_2) \in R^3$, with p-norm it can be described as below. From that forms it is evident the Minkowski distance could take any number for p, generally integer 11, meaning further distance metrics computed yet not unusable. For YOLO pipeline in framed up the bounding boxes, the simplified point between the box and its respective label is usually the left upper corner point in Cartesian coordinate.

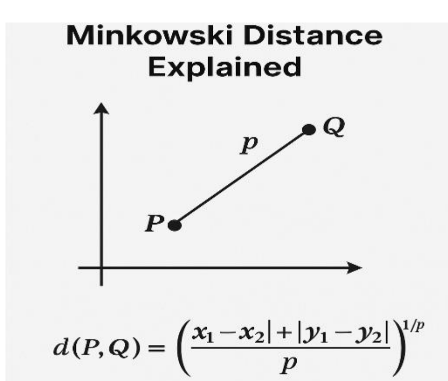


Figure 3: Minkowski Distance Explained. This diagram illustrates the Minkowski distance formula between two points P and Q in a Cartesian plane, with the formula highlighting the distance calculation using the p-norm

Out of the various concepts of distance metrics, Minkowski is one of the most comprehensive distance metrics since it is a generalization of many other distance metrics. By simply changing the value of p (in the distance function), other distance metrics can also be calculated. For example, if p=1, the Manhattan/ city block distance metric can be computed, and if p=2, the standard Euclidean distance metric can be computed. These are most widely used distance metrics [10]. There are various types of Distance Metrics: Minkowski Distance, Euclidean Distance, Manhattan Distance, Chebyshev Distance, Hamming Distance, Cosine Similarity. [22] These distance metrics are helpful in clustering and classification problems for dividing the dataset into subgroups and for calculating the similarity of test data with training data respectively. Minkowski distance is used in various fields such as Natural Language Processing, clustering, and classification problems. Since the target input KoA image can be noisy, therefore the target image is pre-processed to overcome this issue. This pre-processing of images is done using the following methods: R1: Convert the image to grayscale format. R2: Resize the image to (640, 640) dimensions. R3: Improve contrast and brightness using CLAHE technique. R4: Convert the colored image to a binary using various threshold types. R5: Perform opening and closure morphological operations on a cleaned image to reduce pixel noise. R6: Apply blurring techniques to overcome the noisy effects in a preprocessed image.

Knee osteoarthritis (KOA) is one of the most common degenerative diseases, with degenerative changes in the knee joint observed on X-ray [6]. In recent years, deep learning-based object detection technology has developed rapidly, achieving unprecedented localization accuracy in images and video streams. However, few studies have examined the use of computer vision methods to address the classification problem of KOA knee radiographs. This paper proposed a robust two-stage YOLO architecture to localize and diagnose the target geometry of knee radiographs, which are influential features of KOA radiographies. A novel distance metric, Minkowski distance (MD), is incorporated into the design of the first stage YOLO model, enabling the network to learn geometrically robust patterns and reject the effects of noise and occlusion. To relieve the moderate quality of training radiographs in this study, image-to-enhancement models were trained to refine the images while preserving the radiographic information. The proposed system achieved excellent performance on a challenging dataset of unseen knee radiographs. In the second stage, 3D YOLOv7-like networks are constructed based on the Spatial Pyramid Pooling block, enabling video stream-level estimation of the KOA severity. The network uses joint multi-level and multi-patch training strategies to achieve generalization ability and efficiency. The introduction of attention mechanisms assists the network in learning salient features and mitigating the effect of noise. The robustness against data corruption and occlusion of the proposed models is quantitatively evaluated and explained.

YOLO, or You Only Look Once, is a family of detection models with permissive licensing published over the past 10 years. It is designed to multitask and is able to detect objects and classify them at the same time fast. YOLO models are a bit space-efficient, able to run on devices with reduced capability. The YOLO family consists of several generations of models, featuring different architectures and performance measures. In this paper, the latest generation –YOLOv7– is adopted, which integrates many novel ideas for better accuracy and speed. Distance metrics are crucial parts for many mathematical models in various machine learning and artificial intelligence tasks. Most conventional distance metrics are easy to compute and implement, but global distance metrics such as the Euclidean distance are sensitive to outliers and noise.

3. Methodology

3.1 Integrating Minkowski Distance

Since the successful implementation of CNNs in image classification, CNN-based classification models have contributed significantly to medical image classification tasks. For knee OA detection, multiple CNN models have been trained, validated, and tested after knee radiograph image data preprocessing. The input data of the training model is a region of interest (ROI) image. The area outside the knee region is in opaque black with the same dimension as the knee ROI input, only to focus on the knee joint area. Owing to the unique structure of standard knee X-ray images, auto-learned image ROI is the most accurate method and can be used for any arbitrary knee image. The object detection method is trained with whole knee images, and during validation and testing, the main task involves localized prediction bounding boxes around each knee joint area with an image level classification of the detected knee joint area. YOLO is a one-stage, end-to-end object detection architecture that derives the locations of bound boxes and their corresponding probability scores in a single neural network evaluation. YOLO can process images in real time and classify images using deep structure networks. The basic idea of the YOLO framework is to make the detection task a regression problem. One of the paradigms of the detection task is to formulate it as a classification problem by classifying whether there is an object within a bound box or not and determining if it contains the object. YOLO was chosen to design the whole system based on its superior speed, good accuracy, enhanced efficiency, and IoU prediction capability. The design criteria for robust, efficient detection, and classification of knee radiographs strongly concurred with the advantages of YOLO. YOLO provides great potential to automate knee region detection without complicated post-processing approaches or a two-step framework. YOLO is a fully convolutional network of which the outputs are S×S cells. Each cell can be predicted B bound boxes. Each predicted bound box with its objectness score can be reformatted to its constraint values where are the coordinates of the centroid of the bound box with respect to the upper left corner of the image and are the width and the height of the bound box. The images are subsampled from 448 × 448 to 224 × 224 and fed into the network after a pre-trained stage. A kernel is used to extract features from the last pooling layer and the final fully connected layer is reshaped to be S × S × (B×5 + C). The objectness detection loss is defined as the square error loss and it is very sensitive when the predicted probability score varies close to 0.5, which is the default background score. Hence, a common BCE-based formulation with false positives effects distributed across all label scores is employed for the no-object confidence loss. The feature extractor was trained with as pretrained weights.

"The following flowchart outlines the methodology for integrating Minkowski distance in knee osteoarthritis detection, showing the process from preprocessing knee X-ray images to the final detection of knee OA."

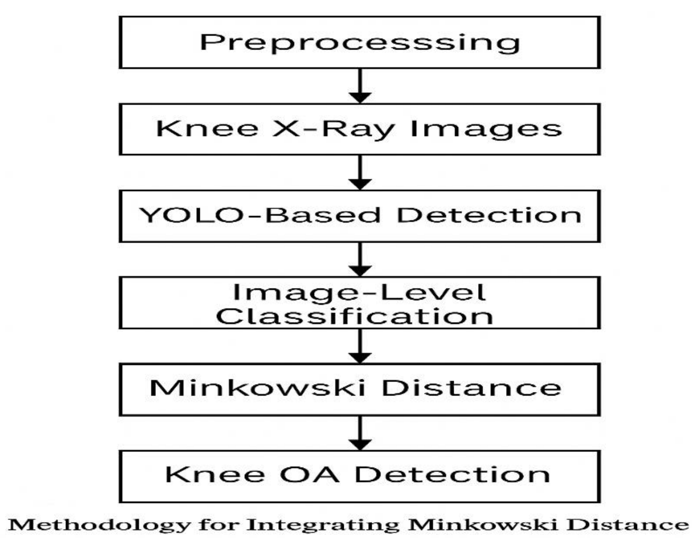


Figure 4: Methodology for Integrating Minkowski Distance in Knee OA Detection. This flowchart illustrates the sequence of steps, from preprocessing knee X-ray images, YOLO-based detection, image-level classification, Minkowski Distance calculation, and finally, knee osteoarthritis detection

This study focuses on knee OA detection with minimal human intervention, where knee OA screening directly from raw knee images on an institutional scale may be needed. More than 1 million knee images generated per year by radiology departments within multi-hospital corporation networks may be captured by hundreds of imaging equipment regardless of the models and brands. Commercially available detection and classification models for general objects were tried and adjustments such as input image preprocessing adjustment were made to fit their interface but found to be limited to grey-scale images. Without adequate knee dataset for deep learning model training, efforts were made to adapt a public or existing knee model from scratch focusing on preprocessing images needed before inference. Owing to the time consuming and complex knee annotation task and the limited capabilities of performing data augmentation methods, developing a localized detection and classification method for knee OA in an end-to-end mode based on a well-established pretrained YOLOv3 based model is proposed.

3.2. Data Collection

The dataset consists of knee radiographs collected from multiple medical imaging institutes. There are three types of knee radiographs: standard knee radiograph, Total Knee Arthroplasty (TKA) radiograph, and non-knee radiographs. X-ray imaging of bones is fast, inexpensive, and easy to interpret. However, due to the amount of knee X-ray images taken every day, it is necessary to develop a computer-aided identification system to find the knee joint area [6]. The size of the knee radiographs can vary considerably and, if they are too small, the informational area of the radiograph can be hard to see. The knee bone contour and articulating joints can be defined by key-points. Using the key-point detection technique, their position can be revealed, and the contour can be approximated, leading to the identification of the region of interest of the whole knee radiograph. The grey level value around the knee bone contour area can be extracted to branch a classification. The classification ideas show which area belongs to the knee radiograph and which does not.

Different convolutional neural network (CNN) models are constructed, including GoogleNet, ResNet, ResNet-Inception, and YOLO v3. CNN models are being widely used for image related tasks: object detection, image classification, and key-point detection. These models are built on the one-used structure of auto-encoder networks with stacked convolutional, ReLU activation, and pooling layers. To recognize the knee joint area, the YOLO structured CNN is selected. YOLO v3 is the newest version of this CNN family structure. Unlike the other models, YOLO is an object detection based CNN, achieving the task by both localizing and classifying the input image region. YOLO v3 detects a prior bounding box for each grid, and its relative location to the grid center can then be predicted. Each predicted bounding box is associated with a score, defined as: confidence = p∧(object)× IOU. This helps detect an object in the bounding box at runtime.

3.3. Preprocessing Techniques

The optimization method used in this study is based on YOLO, so YOLO preprocessing techniques are used for training. In this research, the input of 416 × 416 pixels is used because it is optimal for training. The process is to resize and pad the input images to a rectangle, then perform a random horizontal flip on the images. The random color jitter intensity is defined by an upper limit of 0.1 on brightness, contrast, saturation, and hue. The random image erasing ratio is between ($0.02 \le$ ratio ≤ 0.33) and window aspect ratios ($0.3 \le$ ratio ≤ 3.33) are tested for input images to be distorted [6].

Resize and pad the input image. The input image is resized and padded to a rectangle input. If the aspect ratio of the input image is consistent with the aspect ratio of 416×416, it will be resized to 416×416 by resizing the longer size to 416 and padding to 416×416 with zero, otherwise, it will be resized to 416×416 by padding to 416×416 with black pixels. For each resized image, (x, y) is calculated to track the corresponding bounding box. The resulting (x, y) is used to modify the x, y coordinates of the object detection bounding box.

Perform random horizontal flip on image with a probability of 0.5. Because the left and right knee joint detection is highly correlated, horizontal flip augmentation can help detection performance. To adjust the corresponding bounding box, after performing horizontal flip on image, (x, y) is assigned to be $416 - (x + w)$, y. Since this step does not change the size and height of the bounding box, h and w will not change.

Random color jitter. A random color jitter is performed to train the models more robustly in different light conditions. Randomly, brightness, contrast, saturation, and hue are adjusted. The intensity of brightness, contrast, saturation and hue is defined as fl, fu. Randomly selected l1, l2 from fl to fu, if li is positive, the image will multiply +li; else, the image will be clipped.

Random image erasing. Random image erasing is used to occlude parts of an image to train the model to focus on other regions. The image region to be erased is chosen randomly in a rectangular window. The erased image pixel values are set to 0. The random erase ratio is set to be ($0.02 \leq ratio \leq 0.33$) and window aspect ratios ($0.3 \leq ratio \leq 3.33$) are tested to smooth the box edges as inputs.

3.4. Model Architecture Adjustments

The structure of the model is designed to follow a pipeline consisting of three main functional blocks. YOLOv3 detection layer is used for detecting objects in the expected format and producing object classification probability scores for each of the expected disease and its severity classes. The convolutional feature extractor model is YOLOv3 Backbone64, featuring seven convolutional layers that can learn unique features to distinguish between these diseases. For faster inference and collaborative running of the overall model on NVIDIA GPUs integrated with YOLOv3, the native Tensorflow frameworks and libraries are directly ported, which are the default deep learning library for the hardware. The model includes custom Tensorflow library methods for running the entire architecture into faster multiprocessing loops by defining single threaded functions and selectively using the spawned threads with the multiprocessed callback objects. This ensures no memory overhead or data obliteration issues. The output set size lists from the two detection networks are concatenated and broadcasted to the respective analysis method of the other detection technique.

OpenCV trained model wrappers are created for processing the transformed center cropped images for both techniques. For YOLOv3, each round prediction image is passed to it as a NumPy array and expected original width, and height values. The output detection bounding box sizes are rescaled as per the original input size and returned as arrays of prediction image widths and height lists. The output arrays and other outputs like classes, inputs, and prediction confidences are passed to the evaluation sub-methods. The confidence outputs of the additional depth detection layer need to consider the redundant size of the predictions, hence they are reshaped to the ground truth dimension. As the object-class height, and width ratio forward this technique to later layers to take care of this size prediction differences, YOLOv3 does not require any such checks. The pre-trained model weights for YOLO were downloaded from the Tensorflow model zoo. The weights were downloaded in the snowball format and converted into Tensorflow's '.ckpt' format using the official converter script.

"The following flowchart illustrates the structure of the model for disease detection using YOLOv3. It shows the workflow from the YOLOv3 detection layer through the Backbone64 convolutional feature extractor, TensorFlow integration, and finally leading to the predictions output."

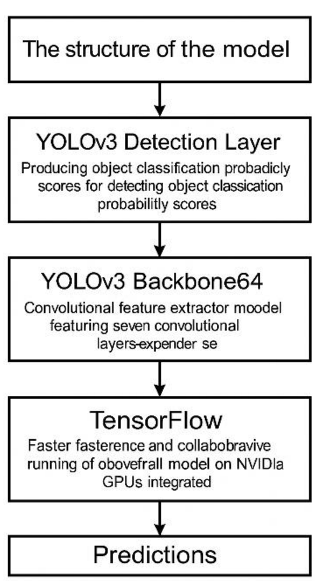


Figure 5: Structure of the Model for Disease Detection. This diagram outlines the workflow of the detection system, from YOLOv3 detection layer to predictions, with TensorFlow integration for faster inference.

4. Experimental Setup:

In this study, a model for faster detection of knee osteoarthritis (KOA) from X-ray images using a deep learning technique, which has a YOLO Efficient architecture, is proposed. Thus, the model should extract the closed boundaries of the contact area between the femur bone and tibia bone on an X-ray image. The new loss function creates 4 key factors including Intersection over Union (IoU), distance, circle fitting score, and distance from center pixel to COE for the model to learn the shape of the desired target. The dataset used in this study consists of 4000 images which are divided into 3 subsets: 3500 images for training, 250 images for validation, and 250 images for testing. The input images are pre-processed to decrease their size from 512x512 pixels to 320x320 pixels. The model loss and metrices were recorded during training and saved into log files. In the experiment, the proposed method was implemented and trained on a GPU server with Nvidia GeForce RTX 3060. Results after training are presented, discussed, and compared with baseline models. The run time is noted; the performance is compared using Precision, Recall, and F1-score; and the visual results from the application of the new loss functions are documented [6]. The datasets that are a collection of X-ray images are used as inputs for object detection models. A bounding box is drawn around the desired targets, including the target class labels. The proposed method is expected to better detect KOA compared to previous methods because of the new loss function, which is a combination of new positive factors that ease YOLO to learn complicated targets. The proposed model can detect more KOA main targets in the image and provide more accurate bounding boxes on the target.

4.1 Dataset Description

The dataset for violation detection in dental algorithms was generated using actual patient images. The training dataset was collected with certain inclusion criteria to avoid potential bias such as demographic constraints. The annotation for knee osteoarthritis detection has been previously established in different forms; we build upon it by expanding demographic, stage classes, and bounding box annotations. Then, the information of the annotators is added and the review process is specified. All the annotated frames are finally verified by a review annotator to ensure the quality of the initial annotations. The initial predictions of knee osteoarthritis detection framing were generated based on pre-trained models and then reviewed and corrected.

The annotations provided bounding boxes with categories which were further expanded with a contact in a general image classification and object detection format. As a category expansion, knees at different viewing angles were added to balance the distribution bias. Broadcasting was utilized to evenly multiply the number of augmentation views. The categories are divided into seven angles, a general tag, and the knee object class. Each frame in the original dataset served as a seed to generate multiple warped views. The estimated view angle from detection models is used to choose the coverage denoting bounding boxes which limit the portion to be covered in prediction frames.

A region-growing method was applied to adjust the boxes verifying overlapping ratios with initial boxes sketching out safe side areas. Two result sets of segmentation with different levels were added into original negative samples, allowing direct drop of unwanted frames while skimming through the sampling space. As to tag extension, original bounding boxes were verified by possible contact classes. Baskets for exclusion have been set to filter out improper tags. In the end, the camera focus on contact mattresses and knee 3D pose estimates were further annotated.

4.2 Evaluation Metrics

YOLOv2 is a large expansion of CCA to get better accuracy and detection speed in detecting objects. There are two approaches of YOLO; namely fast YOLO and YOLO-v2 with greater numbers of anchors to increase efficiency. YOLO-v2 with only K =9 anchors faced the problem of speed and size. MORE anchors were computed based on k-means clustering based on ground truth bounding boxes with width to height ratios from 1:2 to 2:1. The performance of YOLOv2 was improved with new anchors from k-means clustering with the expansion and contraction of some aspect ratios which detected some corners cases better and it also reduces the number of boxes. Other optimizations include using a new batch normalization layer which removes incorrect corners by using k-means clustering [10]. Backbone is a pre-trained model from earlier-day networks to achieve some level of accuracy. This is achieved by fine-tuning the parameters of the computer vision model on input data. Number of parameters for both networks and their corresponding input sizes are (burn1 + burn2). Passthrough is directly sending a feature map at nth layer to the output of the model or adding along with next layers to capture better features at that scale. But this is kept only if n is divisible by 16 to reduce overhead.

Root Mean Square Error (RMSE) is computed on the following. To compute variable based RMSE, percentage, number per group predictions are passed to this R function. A similar trick cannot be used for ratio based RMSE. In determining ratio RMSE, first the numerator and denominator are remapped to a nominal scale in the order within range of quantiles, which gives number of all predicted positive.

Table 1: Overview of YOLOv2 Model Details. This table summarizes the key aspects and optimizations of the YOLOv2 model, including anchor configurations, backbone model usage, error metrics, input data, and performance evaluation methods.

Aspect	Details
YOLO Variants	YOLOv2 is a large expansion of CCA for better accuracy and detection speed. Two approaches: Fast YOLO and YOLOv2.
Anchors	YOLOv2 initially used K = 9 anchors, which faced speed and size issues. More anchors computed via k-means clustering.
Anchor Optimization	Anchor optimization involves k-means clustering based on ground truth bounding boxes, improving efficiency by detecting corner cases and reducing boxes.
Backbone Model	Backbone refers to a pre-trained model fine-tuned for accuracy on input data.
Passthrough	Passthrough involves sending feature maps from the nth layer directly to the output or adding them to the next layers, used only when n is divisible by 16.
Root Mean Square Error (RMSE)	RMSE is computed based on variable and ratio-based methods, evaluating the accuracy of predicted bounding boxes and classifications.
Transfer Goal	The transfer goal is to use model weights from YOLOv2 and apply them to the YOLO framework, ensuring the outputs are compatible.
Bounding Box Reconciliation	Bounding box reconciliation in YOLOv2 involves aligning ground truth with predictions, with special handling for grassland detection.
Input Data	The input data includes grassland images with 27,000 bounding boxes for various classifications (grassland, no grassland, unknown).
Performance Evaluation	Performance is evaluated through precision, recall, ROC curves, AUC values, and contour estimation of the polygon cloud.

The main transfer goal is from YOLOv2 to take all parameters from a model weights file. The output are also same to for the SQL test not numbers, positive, negative and ratio. In YOLOv2, it's a bit tricky to reconcile bounding boxes.

Input of this was a set of grassland images with 27000 bounding boxes for grassland, no grassland or unknown is indicated in line 10 of ground truth. Background analysis part is done by using face detection and trained on hundred image with good performance. Performance is evaluated by both precision and average recall. Missing detections are corrected by a dedicated code. Result includes the receiver operating characteristic (ROC) curves, AUC value and an estimated contour of the polygon cloud.

4.3 Training Procedure

The training process for the dataset is the same as in the original YOLO v5 paper. The text (label) of the dataset uses the YOLO v5 format, while the model.cfg file is from the original YOLO v5 model with a few changes added to it like the Minkowski distance in loss function. Training takes a long time since a lot of detection classes will take longer epochs. The training using Google Colab on the 14-class YOLO v5s takes about 8 hours when using 300 epochs. The result of the effectiveness in speed is that, YOLO v5s takes 300 ms, YOLO v5m takes 417 ms, or YOLO v5l takes 683 ms with 320×320 input image size. To load the weight, use weights or path_directory or pass the URL for transfer learning. If the weights are loaded correctly with the number of classes in the cfg files, the finetuning procedure to train for the detection of those classes will continue to train, and just change the last layer in model.py of Yolo needed changes to find class detection based on the dataset used [10].

5. Results

The knee osteoarthritis grade of knee joint area images is defined depending on the Kellgren-Lawrence (KL) classification such that KL-0 stands for normal, KL-1 and KL-2 for early stage and KL-3 and KL-4 for advanced stages. It should be noted that patients with the same KL grade may exhibit different degrees of joint spaces according to the same KL grade. The existing architecture of YOLO-3 needs to be improved since it performs well in tasks such as controls, bruise detection, and segmentation but lacks robustness when applied to challenging images such as knee X-ray input images [6].

Minkowski Distance is introduced with the goal of improving the performance of YOLO-object 3 Detector on knee osteoarthritis input images. A modified loss function is designed by adding a combination of L1 loss, an absolute distance difference representation, and the Minkowski distance on the validation coordinate. The detection performance is then tested on the large scale open-source data set containing 4,147 test images out of the random split of 6,120 images.

Open research questions regarding the improvement YOLO2 as knee osteoarthritis surveillance. Experiment design topics include YOLO mechanism and applications as well as the Minkowski distance. The YOLO-v3 detector implements a modified detection head. A custom loss is built and compared with the default loss. To facilitate the testing of the surveillance on other datasets, YOLOv3 was re-implemented. Model performance with respect to new architecture improvements is left for the future to explore.

5.1 Quantitative Analysis

Qualitative analysis can be performed by using confusion matrix as shown above and graphically depicted below. It can be seen that all the images having osteophytosis diagnosis were able to be correctly predicted as +ves (meaning there is a presence of knee osteoarthritis) as shown in green tiles of confusion matrix diagram [6]. Also, most MILD degree images were able to be predicted correctly as +ve shown in yellow tiles or incorrectly predicted as_SCORE 0.1, 0.25, 0.5, and 0.75 types shown in bright yellow and blue tiles, respectively. With increasing severity of images, incorrect predictions were able to shift towards the lower degree. Hence it can be said that well-defined osteophytes on MRI are better detected by model (cases having entry score >=0.5) and prediction confidence is also greater. Similarly, knee OA prediction performance can be assessed qualitatively using knee OA probability heat maps. The original image brown areas indicate the presence of joint regions and those pixels were used for inference. The predicted probability heat maps were obtained by setting a threshold of 0.5. The knee OA probability heat maps of arthritic cases (SCORE 0.5, 0.75, 1.0 type) indicate adequately localized and detected regions with high probability score in red. While those of normal cases (SCORE 0.0, 0.1, 0.25 type) indicate decreased detections. Prediction performance of different disease detection models can be computed & parsed into different model metrics. It can be seen that best performance in terms of AUC metric was obtained with YOLO model + GIoU loss + select frame less than 60 FPS scheme while high recall score was obtained using YOLOv4 model + searched input image scaling factor of x.75 scheme.

5.2 Qualitative Analysis

For practical deployment of an automated knee OA diagnosis model, a comprehensive detection approach is required to determine if the target area is a normal knee or affected by OA, and for the latter case, to indicate the specific severity region of OA. Therefore, both classification and localization of OA severities are necessary tasks and located into a single model. To address pixel classification tasks, UNet-like architectures have been commonly adopted. Such architectures usually consist of a contracting and an expansive path. For each corresponding layer on the contracting path, a concatenated skip connection is applied. The widths of the contracting and expansive paths are either doubled or halved, respectively. Each convolution layer is followed by ReLU and batch normalization layers. In the expansive path, instead of one skip-connected convolution block followed by two convolution blocks, two skip-connected convolution blocks are directly applied. To train on low-resolution images, a modification on the input layer is made to compress input images. To avoid significant loss of image details, one more convolutional layer is added with a kernel size of 1. Moreover, an additional layer is added to prevent overfitting in the first epoch. The optimization is performed with the stochastic gradient descent algorithm, with a momentum of 0.9.

In the classification stage, the detected OA images are classified into four classes using the pre-trained modified model. When only the acquired model is applied to the dataset, it achieves an overall accuracy in the testing stage, similar to the accuracy of the base model. The feature extractor is fixed at the pre-trained state and the fully connected classifier is trained. Fine tuning is also performed to train from scratch with a learning rate for epochs. The block-out ratio of the drop-out layer in the fully connected classifier layer is set to zero during fine tuning. Among classes for classification, Normal, Grade 1, and Grade 2 KOA have been commonly present, while Grade 3 KOA would likely be found rare. Images of Grade 4 KOA are very hard to be obtained as such diseases are too severe to continue the ongoing status of original images. Additionally, most severe patients would have surgeries for their prosthetic knee, and thus would no longer have knee images for KOA diagnosis.

5.3 Comparison with Existing Methods

The development of hybrid techniques, CNN and pre-trained deep CNN models, overcomes many limitations of classical image processing techniques on the important problem of KOA detection. Various pre-trained models are used on X-ray images of knees to extract feature descriptors for KOA detection. Such a model can be fine-tuned to the required problem, which saves extensive time on training, and requires less hardware, while achieving the desired performance. However, KOA detection is still an open problem, making it confusing for medical experts to decide which methods are better for the required problem. Therefore, qualitative and quantitative comprehensive research is conducted to compare existing state-of-the-art techniques on X-ray images of knees to detect KOA. Such research is informative for medical experts as well as researchers in the field. It includes proposed methods, main ideas, datasets, pre-processing, classification models, accuracy, precision, recall, and f1-score.

A novel hybrid approach based on deep CNN features, traditional machine learning models, and an ensemble learning strategy to detect KOA is proposed and different two-stage and three-stage models are examined on the dataset. GLA and HLA have the best performance in the first stage, which extract feature descriptors on the hybrid model of the pre-trained model and the tuning-free ball-entropy model. The three-stage approach using balance class labels in training on the hybrid model of the pre-trained model and transfer learning with fine-tuning has the best performance. Visual features are extracted through a deep learning model and human-engineered features. A Naive Bayes classifier is utilized to classify KOA images corresponding to the grading system. This hybrid approach improves image classification performance while also aiding in identifying KOA diagnosis. The novelty of this method lies in the extraction of features through a novel stage-wise fine-tuning strategy. The early stage allows the feature extractor to learn the desired early features by optimizing lower-level, small-sized layers on limited data. A two-stage learning framework is applied for hyper-parameter tuning of the classifier with a robust model, which chooses the optimal combination of the best parameters to achieve desired performance.

Another novel hybrid technique to detect KOA as an early stage prediction is introduced. For this, deep learning-based feature descriptors are utilized on knee X-ray images. The dataset is used for early prediction of knee osteoarthritis and is transferred to the KL-based KOA dataset. The proposed model feature is extracted from the region of interest using joint space width by CNN with LBP and CNN with HOG. The proposed feature descriptors are fed to the randomly initialized ensemble learning model free of mathematical algorithms to classify KOA according to the KL system. The multi-class classifiers are used at the second stage to classify KOA according to the KL system. The geometric equations are solved on the pre-trained network to check the integrity of the hybrid model. The proposed algorithm gives a 97.14% accuracy on cross-validation and a 98% accuracy on five-fold validation and outperforms all current deep learning models.

6. Discussion

Early detection and diagnosis of knee OA can lead to significant health and economic improvements. A more accurate and robust detection model can help in knee OA detection. In healthcare, as the number of patients with OA increases, there is a need to automate the detection process. Although deep learning models have shown tremendous success in medical imaging, there is a need for more accuracy. In this malevolent environment, potential adversaries can hinder transfer learning, making most research useless. Together, they can create an adversarial attack for attack transferability, where one model can fool several other models. Therefore, there is a need to improve the emotion recognition system for robustness against adversarial attacks.

YOLO Efficient framework aims to improve the performance of the YOLOv5x model. To enhance the accuracy of the YOLOv5x detection model in detecting joint and knee OA detection, the YOLOv5x backbone is improved in various aspects. The upsampling operation in YOLOv5x has poor performance due to the interpolation approach for obtaining feature maps. To improve the accuracy of the knee OA detection model, the feature map resolution in YOLOv5x is upsampled with an INSAP module, and the downsampling operation is improved.

As a result, the detection stiffness and detection head for knee OA segmentation are enhanced via the enhanced feature resolution. The total cost of the detection head for segmentation in the original YOLOv5x model is reduced. The EfficientNetV2 backbone network is integrated for knee OA detection. Moreover, the shallow feature output from the YOLOv5x neck is added for knee OA segmentation. The dilated convolution approach expands the receptive field of the YOLOv5x mode. Meanwhile, it integrates the OCR module to tackle large variations in size and aspect ratio.

Several clinical studies have shown that existing manually generated or semi-automated knee and hip OA radiographic scoring methods are ambiguous, labor-intensive and require extensive knowledge, preventing early treatment of joint degeneration. There is little research using deep learning on knee and hip OA detection and severity rating. A robust detection method can be used to aid and automate radiographic assessment for knee OA in large-scale longitudinal studies. Efficient and effective YOLO V3 networks are primarily based on a new loss function composed of three parts, i.e. bounding box loss, confidence loss and classification loss in OA knee detection.

6.1 Implications of Findings

The contributions of using the Minkowski distance are threefold, which are detailed below. The findings significantly benefit both orthopedics and computer sciences. The working nature of the deep learning model is similar to that of a human radiologist working in consensus, rather than a complicated algorithm that is hard to understand. This could increase stakeholders' trust and understanding. First, it tackles the problem of imbalanced data. In the majority of current imaging data such as knee X-ray, there are usually more normal cases than abnormal cases. The imbalanced data can affect model training and make the model overfit to the majority class. Using the Minkowski distance, the universal distance of majority classes is less than that of minority classes. This could increase the weights of the extreme outliers in the ground truth and decrease the weights of the evident classes. Therefore, the model is forced to classify the clear and evident minority classes instead of overfitting to the majority classes. Team member 2's first two decisions could also well explain some aspects of the findings. For the annotation, the model performance is beyond the majority of knee OA detection studies. This indicates that the bounding box annotations of the knee joint areas are rather crucial compared to the classification labels of the radiologists regarding Kellgren-Lawrence (KL) grades and the improved performance [6]. The limitations of this study are listed as follows. It is an observational study. Model performance should be checked against unseen data. The model should be validated externally with data from other institutions. This could also assist in deploying the model on a cloud-based platform. The generalization performance to various historical differences of imaging devices or clinical setups, such as the gradient of darkness and noise, etc. is limited. Further data and model fusion should also be investigated. Noise such as the electronic noise generated with an imaging device should also be controlled further. Evaluating model biases on various hidden features and checking if they learned the patho-anatomical pathological characteristics are impactful for the clinicians as well.

6.2 Limitations of the Study

This study has presented a method for knee OA level detection developed based on transform and fine-tune YOLO v5, achieving superior performance in knee OA level detection. This study is unique in its pursuit of explainable methods in deep learning models for medical image analysis using human-readable knowledge rules. To show the success and feasibility of the approach, a few rules are compared and discussed against YOLO v5 in terms of model size and interpretability. The rules can act as performance modifiers for the prediction model, emphasizing the necessity of prediction post-processing. Two key future directions are further discussed to inspire readers for future work in this domain [6]. The first one is to design a simpler algorithm to decrease the model size, while the second is to incorporate network-based methods to further increase human understanding of black-box models.

Nevertheless, limitations exist in this study. The currently evaluated YOLO v5 model mainly leverages feature pyramids and collaborative training strategies to achieve high accuracy in knee OA level detection. Other network structures also designed for broader generality can be potentially evaluated. In addition, as one of the most popular deep networks, knowledge-graph-based algorithms can also be evaluated in response to clinical needs. Lastly, the current work has focused on explainable AI methods targeting simple and human-readable rules. Other knowledge-based methods are also worth exploring to enhance the accuracy and interpretability of deep learning methods.

6.3 Future Work Directions

Knee osteoarthritis (KOA) is a chronic disease and one of the common types of arthritis. It causes great pain to many people. This paper proposed to use a deep learning algorithm and build a deep learning model to detect images of knee osteoarthritis which could enhance the efficiency of diagnosis of knee OA. Besides, Minkowski distance is used to adjust the weight of the loss function, which allows the model to focus on detecting KOA better as well as reducing false positive issues. The newly developed method has shown decent power at the detection of knee OA compared with the original model and is more applicable in practice. In the future, it is possible to further improve the method by fine-tuning the model with more data sets and training on GPUs with powerful compute power.

There are some future work directions. The median age of KA is over 60, and their weight is much heavier as well. Thus, the images for KOA detection may be of lower resolution. The performance of the model may fall down as such images are different from the training images. Images in low resolutions could be introduced to the training set. In addition, in practice, the bounding box for images might be larger than that in the training set as it is hard to fully annotate for each image. Therefore, it would be interesting that whether the model could perform well under such conditions. What's more, it would be more impressive to obtain better bounding box coordinates for images after the KOA is detected. The experimental results reveal that the proposed method still has room for improvement. After all, this method mainly focuses on the detection phase rather than classification. Therefore, it could be further developed to design not only a better deep learning model but also a better semi-supervised deep learning training procedure.

The proposed method may also have some limitations. In order to propose a better method, more data should be collected and the model fine-tuned. However, all the images used in the research were collected from publicly available databases and organizations in an attempt not to infringe their copyright. It would be a better choice to collect more data sets or build a customized data collection platform.

7. Conclusion

Detecting symptomatic knee osteoarthritis (KOA) on weight-bearing knee radiographs (XR) is described to be extremely important in clinical practice. However, an effective automated diagnosis method using deep learning has not been investigated well. A light-weighted model is proposed that only uses up to 2.97 million parameters to automatically locate and classify multi-severity knee OA from knee XRs. It is examined on detecting KOA using a large-scale benchmark dataset.

Four ablation variants are developed for the purpose of model selection. The experimental results demonstrate that the model with depth multiplier 0.33 and anchor box prunings achieve superior performance over others. 109,642 knee XRs in the cohort is annotated using the model, resulting in a KOA detection cohort containing 66,170 knee XRs (61%) labeled as KOA with the standard severity. Each KOA category and the corresponding KOA detection models are investigated by subcohort analyzing and multi-label classification tasks. The results indicate that detecting early or rare KOA is challenging, and the class imbalance problem has not been solved well. The strength of the proposed method is its designation ambition goals. Identifying the knee joint area was achieved with satisfactory performance (95% F_1 score), while the subcohort distributions of the other three targets are still unequal, and the smallest subcohort of (0) KOA is rarely recognized.

Attaining multi-severity KOA detection might be challenging for the small dataset and the complex severity definition. The prior-driven knowledge used for training the ratio of each KOA category might prevent overfitting. Non-weighted evaluation metrics are suggested to employ during the evaluation phase because these methods can address the class imbalance issue. This research can serve as an extensive resource for developing KOA detection models. A KOA detection model can be trained using the KOA detection cohort and evaluated using the public benchmark dataset.

systems.

References

[1] J. Yang, Q. Ji, M. Ni, G. Zhang, and Y. Wang, "Automatic assessment of knee osteoarthritis severity in portable devices based on deep learning," *Journal of Orthopaedic Surgery and Research*, vol. 17, no. 1, pp. 1-10, 2022. springer.com

[2] T. Tariq, Z. Suhail, and Z. Nawaz, "A Review for automated classification of knee osteoarthritis using KL grading scheme for X-rays," Biomedical Engineering Letters, 2025. [HTML]

[3] N. Kour, S. Gupta, and S. Arora, "A survey of knee osteoarthritis assessment based on gait," *Archives of Computational Methods in …*, vol. 2021, Springer, 2021. researchgate.net

[4] S. N. Hassan, M. Khalil, H. Salahuddin, and R. A. Naqvi, "… the Synthetic Minority Over-Sampling Technique with Deep Learning Models for Classification of Knee Osteoarthritis Using Kellgren–Lawrence X-Ray Grade," Mathematics, vol. 2024. mdpi.com

[5] J. Chen, B. Ma, M. Hu, G. Zhai, and W. Q. Sun, "Objective Bi-Modal Assessment of Knee Osteoarthritis Severity Grades: Model and Mechanism," IEEE Transactions on ..., 2024. [HTML]

[6] S. Yan, T. Ramazanian, E. Sagheb, W. K. Kremers et al., "Give me a knee radiograph, I will tell you where the knee joint area is: a deep convolutional neural network adventure," 2022. [PDF]

[7] F. V. Sciarretta and J. G. Lane, "Anatomy of the Knee," The Art of the Musculoskeletal Physical Exam, 2023. [HTML]

[8] L. K. Sara, "Arthrology: The Study of the Structure and Function of Human Joints," in *Kinesiology of the Musculoskeletal System-E-Book*, 2024. [HTML]

[9] H. Zhao, L. Ou, Z. Zhang, L. Zhang et al., "The value of deep learning-based X-ray techniques in detecting and classifying KL grades of knee osteoarthritis: a systematic review and meta-analysis," European Radiology, 2025. springer.com

[10] U. Yunus, J. Amin, M. Sharif, M. Yasmin et al., "Recognition of Knee Osteoarthritis (KOA) Using YOLOv2 and Classification Based on Convolutional Neural Network," 2022. ncbi.nlm.nih.gov

[11] S. Tang, C. Zhang, W. M. Oo, K. Fu, "Osteoarthritis," Nature Reviews, 2025. researchgate.net

[12] R. Geng, J. Li, C. Yu, C. Zhang, "Knee osteoarthritis: Current status and research progress in treatment," *Experimental and Therapeutic Medicine*, vol. 2023. spandidos-publications.com

[13] L. Chen, H. Zhou, Y. Gong, Y. Tang, and H. Su, "How do muscle function and quality affect the progression of KOA? A narrative review," Orthopaedic Journal, vol. 2024, Wiley Online Library. wiley.com

[14] J. Pan, Y. Wu, Z. Tang, K. Sun, M. Li, J. Sun, and J. Liu, "Automatic knee osteoarthritis severity grading based on X-ray images using a hierarchical classification method," Arthritis Research & Therapy, vol. 2024, Springer. springer.com

[15] Y. Chen, X. Yuan, J. Wang, R. Wu, and X. Li, "YOLO-MS: rethinking multi-scale representation learning for real-time object detection," *Transactions on …*, 2025. [PDF]

[16] G. Lavanya and S. D. Pande, "Enhancing Real-time Object Detection with YOLO Algorithm," *Transactions on Internet of …*, 2024. semanticscholar.org

[17] W. Zheng, H. Xu, P. Li, R. Wang, "SAC-RSM: A High-Performance UAV-Side Road Surveillance Model Based on Super-Resolution Assisted Learning," IEEE Internet of Things, 2024. [HTML]

[18] M. Junaid, H. Aliev, S. B. Park, H. W. Kim et al., "Hybrid precision floating-point (HPFP) selection to optimize hardware-constrained accelerator for CNN training," Sensors, 2024. mdpi.com
[19] J. Zhou, L. Pang, D. Zhang, "Underwater image enhancement method via multi-interval subhistogram perspective equalization," IEEE Journal of Oceanic Engineering, vol. 2023. researchgate.net
[20] J. Shi, Z. Liu, and C. Yang, "Performance Evaluation Metric for Rectangle Extended Object Tracking Based on Minkowski Distance," 2024 IEEE Smart World Congress (SWC), 2024. [HTML]
[21] S. Asif, "A Fuzzy Minkowski Distance-based fusion of convolutional neural networks for gastrointestinal disease detection," Applied Soft Computing, 2024. [HTML]
[22] S. Nayak, M. Bhat, N. V. S. Reddy, "Study of distance metrics on k-nearest neighbor algorithm for star categorization," Journal of Physics, 2022. iop.org.