Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-Time Osteoporosis Diagnosis from Knee X-Rays Using anEnhanced YOLO Architecture on Jetson Nano

SUDHIR SANGULA
VIT-AP University
Dr.Siddique Ibrahim S.P.
siddique.ibrahim@vitap.ac.in

VIT-AP University

Research Article

Keywords:

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8446858/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Journal Name

Crossmark

RECEIVED
dd Month yyyy

REVISED
dd Month yyyy

PAPER

Real-Time Osteoporosis Diagnosis from Knee X-Rays Using an Enhanced YOLO Architecture on Jetson Nano

Sudhir Kumar Sangula[1], Dr.Siddique Ibrahim S.P.[2]

[1]School of Computer Science and Engineering, VIT-AP University, Amaravathi, India
[2]School of Computer Science and Engineering, VIT-AP University, Amaravathi, India
*Author to whom any correspondence should be addressed.

E-mail: sangvi.sudhir@gmail.com ,siddique.ibrahim@vitap.ac.in

Keywords: Osteoporosis Detection, Knee X-ray Analysis, C3K2 Module, A2C2F Module, Jetson Nano board, Medical Image Analysis

Abstract

Automated osteoporosis detection from knee X-rays poses several challenges, including subtle fracture patterns, poor feature representation across scales, and misclassification due to overlapping bone structures. To overcome these limitations, we propose a novel YOLO-based architecture incorporating three key modules: (i) the C3K2 module, which efficiently downsamples while maintaining robust gradient flow through dual-path processing, (ii) the A2C2F module, which enhances feature learning by splitting inputs into parallel convolution branches before fusion, and (iii) channel-wise concatenation, which merges multi-scale features to improve diversity and spatial coherence without information loss. Experimental evaluation shows that Our model achieves state-of-the-art performance, attaining an mAP@50 of 0.991 for three classes dataset, and mAP50 of 0.892 for two classes dataset, outperforming benchmarks such as YOLO versions. Furthermore, we deployed the model on an NVIDIA Jetson Nano, demonstrating efficient real-time inference with accurate classification of Normal and Osteoporosis cases. This combination of architectural innovations and hardware deployment highlights the model's potential as a portable, reliable solution for osteoporosis diagnosis, particularly in low-resource clinical environments.

1 Introduction

Osteoporosis is a chronic skeletal disease marked by reduced bone mass and deterioration of bone microarchitecture, resulting in increased fragility and a higher risk of fractures. Among the various affected skeletal regions, the knee joint is of particular concern because it plays a crucial role in weight-bearing and mobility. Knee osteoporosis often develops due to the progressive thinning of trabecular and cortical bone surrounding the knee joint, especially the distal femur and proximal tibia, which compromises joint stability and functional strength. The prevalence of knee osteoporosis has been increasing globally, particularly among the elderly population, postmenopausal women, and individuals with sedentary lifestyles. According to clinical findings, knee osteoporosis frequently coexists with degenerative conditions such as osteoarthritis, complicating diagnosis and management. Patients suffering from knee osteoporosis commonly experience symptoms such as chronic knee pain, reduced joint strength, limited mobility, and an increased risk of fractures even from minor falls or impacts.

Early detection and diagnosis of knee osteoporosis are critical because the disease progresses silently in its initial stages, with symptoms becoming evident only after significant bone loss has occurred. However, conventional methods have limitations in terms of sensitivity, cost, and accessibility, particularly in resource-limited settings. Recent advances in medical imaging and computational intelligence, including machine learning and deep learning approaches, have shown promising potential in improving the early diagnosis and classification of osteoporosis. Automated image analysis systems can identify subtle bone texture changes and microstructural patterns in knee radiographs that may be overlooked in manual assessments. These AI-driven diagnostic models not only enhance accuracy but also provide scalable and cost-effective solutions for large-scale population screening.

2 Literature Survey

The advances in artificial intelligence and deep learning have revolutionized the study of knee osteoarthritis (KOA) research has expanded considerably in recent years, addressing the challenges of subjective interpretation and limited reproducibility in traditional radiographic grading systems. Early work by Thomas et al.[1] presented an automated classification framework using deep neural networks for radiographic KOA severity assessment. Their study highlighted how AI-driven models could achieve consistent performance compared to radiologists, thus establishing a foundation for subsequent developments in automated grading. Object detection methods also gained traction for KOA quantification. Liu et al.[2] employed an improved Faster R-CNN model to automatically quantify KOA severity. By adapting a widely used computer vision architecture to medical imaging, they demonstrated that region-based deep learning methods can successfully capture localized features associated with disease severity. Building upon CNN architectures, Zhang et al.[3] proposed an attention-based deep learning model for Kellgren–Lawrence (KL) grade classification using Osteoarthritis Initiative (OAI) data. Their use of attention mechanisms allowed the network to focus on discriminative image regions, thereby improving accuracy in KL grading.

Alternative architectures, such as densely connected convolutional networks, were introduced by Mikhaylichenko and Demyanenko [4]. Their approach aimed at capturing fine-grained radiographic features for automatic KOA grading, demonstrating the value of deep connectivity in enhancing feature reuse. Similarly, Leung et al.[5] extended KOA classification to prognosis by predicting total knee replacement alongside diagnosis, using deep learning models applied to large-scale radiographic datasets. This study provided evidence that AI could not only classify disease severity but also support longitudinal outcome prediction. Multimodal feature integration has also emerged as a promising direction. Kim et al. [6] investigated whether incorporating additional patient information, such as demographic and clinical data. Their findings confirmed that multimodal approaches outperform image-only models, suggesting that a combination of imaging and patient-specific variables is critical for accurate severity estimation. Tiulpin and Saarakkala [7] contributed by developing CNN models for grading individual KOA features from radiographs, thereby offering a more detailed clinical interpretation than global severity grading.

Although most studies were KOA-focused, broader applications in medical imaging enriched the methodological landscape. Chaves et al. [8] evaluated transfer learning with pre-trained CNNs for breast cancer detection in infrared imaging. Their findings demonstrated that transfer learning can effectively adapt models to new domains, reinforcing its potential applicability to KOA diagnosis where labeled data is often limited. As research progressed into 2021, Yong et al.[9] addressed a fundamental issue in KOA grading by introducing an ordinal regression module for severity classification. Their model captured the ordered nature of KL grades, thereby reducing inconsistencies in categorical classification. Saini et al.[10] carried out a comparative analysis of existing classification and grading methods, systematically evaluating performance across models and datasets. This benchmark study provided critical insights into the strengths and limitations of state-of-the-art approaches. Beyond imaging, Badshah et al. [11] explored genetic markers for early diagnosis of osteoarthritis in a Pakistani population. Their study emphasized the role of genetics in disease susceptibility and suggested that combining genetic markers with radiographic imaging could lead to more personalized prediction models. Methodologically, Shi et al.[12] proposed a deep neural network for rank-consistent ordinal regression leveraging conditional probabilities, directly tackling inconsistencies in ordinal predictions and further refining the accuracy of severity grading models.

Recent research has also focused on reviewing and optimizing the combination of imaging and machine learning approaches. Teoh et al. [13] conducted a comprehensive review of manual imaging grading techniques and machine learning methods for KOA diagnosis and prognosis, highlighting both clinical challenges and technological opportunities. On the methodological front, Liu et al.[14] introduced a novel focal ordinal loss function designed to enhance model performance in class-imbalanced datasets. Their approach enhanced robustness by penalizing misclassifications more effectively, an important step toward clinical applicability. The most recent contributions have explored broader machine learning frameworks. Tariq et al.[15] proposed approaches for KOA classification using machine learning techniques, further demonstrating the diversity of computational methods being applied in this domain. Their work reflects the growing interest in hybrid and ensemble models to maximize prediction performance.

The Key contributions of this work include

- We designed a novel **C3K2 module**, an enhanced cross-stage partial network with a kernel size of 2, which effectively down samples the feature maps while maintaining gradient flow. This module splits the input into two paths, processes them through a series of convolutions,

and then concatenates the outputs, resulting in a computationally efficient and feature-rich representation.

- We introduced an **A2C2F module**, an advanced bottleneck structure that splits the input into multiple branches, processes each through a sequence of convolution layers, and then fuses the features via concatenation and a final convolution. This design promotes better feature flow and representation learning, leading to improved model performance.
- We utilized a **concatenation** operation that merges feature maps from different network paths along the channel dimension, thereby increasing the feature diversity and improving the model's ability to capture multi-scale information without losing spatial coherence

3 Proposed Model

Fig.1. describes three main sections:

1.Backbone (Top Section): A series of convolution layers (Conv, C3K2) that extract hierarchical features from the input image. This is the feature extractor.

2.Neck: It uses upsampling and concatenation to combine features from multiple backbone levels, enabling the model to effectively detect objects of varying sizes.

3.Head: The final part that takes the processed features from the neck and produces the actual detections (bounding box coordinates, objectness score, and class probabilities).

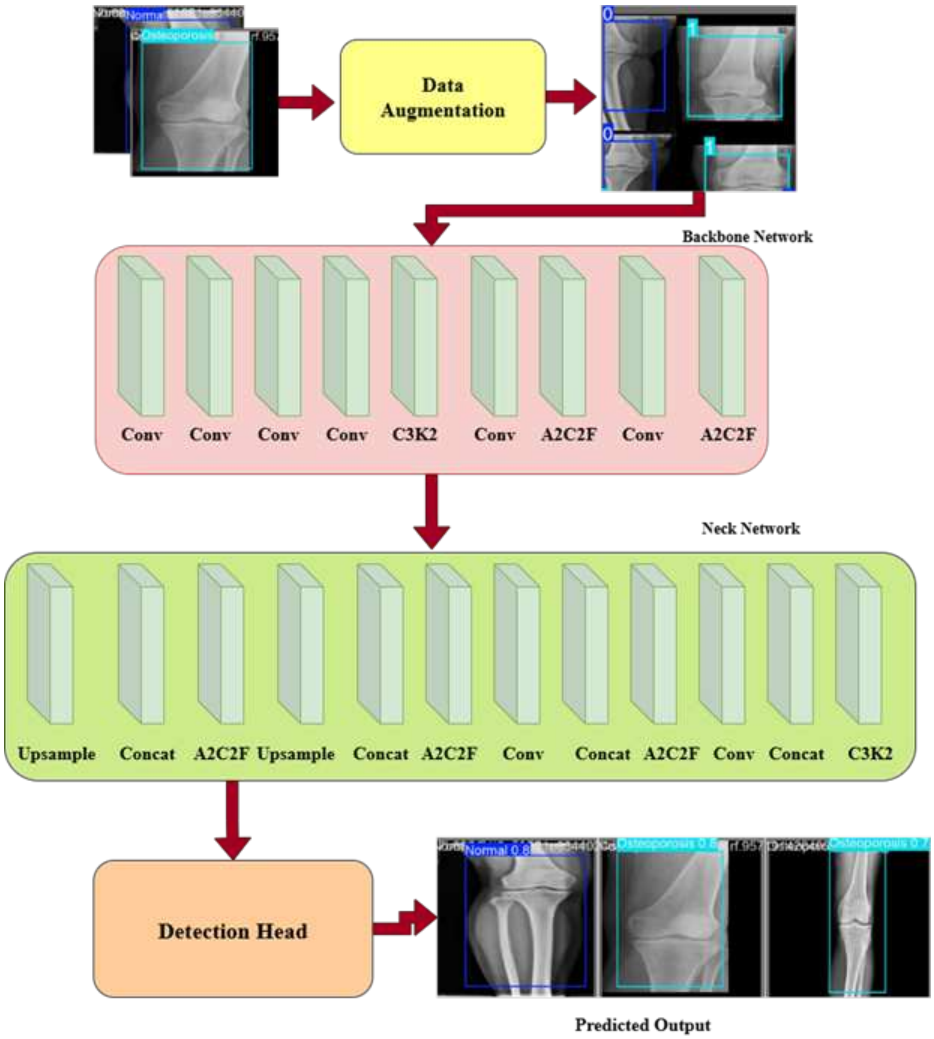


Figure 1: **proposed model layer architecture.**

3.1 ***Backbone Blocks***

A typical convolutional (Conv) block consists of several components: first, a *convolution operation*, where a set of filters (kernels) slides over the input to generate feature maps; second, *Batch Normalization (BN)*, which normalizes the outputs of the convolution to stabilize and accelerate training;

and third, an *activation function* that introduces non-linearity. The convolution operation produces an output $Z = W * X$ with dimensions $(C_{out}, H_{out}, W_{out})$, representing the extracted feature maps for the next layer. In a standard convolution block, the output of the convolution is normalized using Batch Normalization. This process stabilizes and accelerates training by subtracting the mean (μ_B) and dividing by the standard deviation $(\sigma_B^2 + \varepsilon)$ calculated across the current mini-batch, then scaling and shifting the result using two learnable parameters, γ and β. This normalized output is then passed through an activation function, such as the SiLU (or Swish), which is simply the product of the normalized value and its sigmoid $\sigma(\cdot)$, introducing a smooth, non-linear gating mechanism that helps the model learn more complex patterns.

$$\hat{Z} = \frac{Z - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{1}$$

$$BN(Z) = \gamma \hat{Z} + \beta \tag{2}$$

C3K2 (Cross Stage Partial Network with Kernel Size) is a more complex residual block. The "C3" is a term from YOLOv5/v8, representing a bottleneck structure with 3 convolutional layers. "K2" likely refers to a kernel size of 2 for one of the convolutions (e.g., for downsampling) or a stride of 2. A standard C3 block splits its input into two paths, processes them through several convolutions, and then concatenates the results. It's designed to be computationally efficient while maintaining a rich gradient flow. Simplified Mathematical Flow:

1. **Split:** Input X is split into two parts: X_1 and X_2.
2. **Path 1 (X_1):** Passed through one or more convolutional (Conv) blocks.
3. **Path 2 (X_2):** Passed through a series of Conv blocks (the "bottleneck").
4. **Concat:** The outputs of both paths are concatenated:

$$Y = \left[\text{ConvPath1}(X_1),\ \text{ConvPath2}(X_2)\right] \tag{3}$$

Final Conv: The concatenated result YY is passed through a final Conv block to fuse the features.

3.2 ***Neck Block***

Its goal is to mix high-resolution, low-semantic features (from early layers) with low-resolution, high-semantic features (from deeper layers). Upsample block increases the spatial dimensions (Height and Width) of a feature map. This is typically done using Nearest Neighbor or Transposed Convolution (Deconvolution). Nearest Neighbor (Common): Each value in the input is simply repeated to create a larger map. If the upsampling scale factor is 2, one pixel becomes a 2x2 block of the same value.

$$Y_{2i,2j} = Y_{2i+1,2j} = Y_{2i,2j+1} = Y_{2i+1,2j+1} = X_{i,j} \tag{4}$$

The concatenation (Concat) operation merges two feature maps by combining their channels, requiring that their spatial dimensions (height H and width W) are identical. It takes two input tensors, A with shape (C_A, H, W) and B with shape (C_B, H, W), and produces an output tensor Y = [A; B] by stacking their channels together, resulting in a new combined feature map with shape $(C_A + C_B, H, W)$.

The A2C2F block is a specialized, efficient module likely inspired by designs like the C2f block in YOLOv8. Its function is to enhance gradient flow and feature representation through a multi-branch bottleneck structure. The process begins with an initial convolutional layer that projects the input features to a specific channel size. This output is then split into several parts; one part may bypass further processing while the others are fed through a sequence of two or more convolutional layers (the "bottleneck"). Finally, all parts are concatenated along the channel dimension, and a final convolutional layer fuses this combined information, allowing the block to capture and refine complex features from different receptive fields effectively.

3.3 ***Detection Head***

This is not detailed in the diagram but is standard for one-stage detectors like YOLO. Its purpose is to make the final predictions from the feature maps provided by the neck.

Mathematical Operations:

1. **Input:** Multiple feature maps from the PANet (e.g., at three different scales for small, medium, and large objects).

2. **Convolutional Layers:** A few final Conv layers to adjust the number of channels to the required output depth.

3. **Output Tensor:** For each spatial cell in the feature map, the head predicts:t_x, t_y, t_w, t_h: Offsets for the bounding box center and dimensions.p_o: An *objectness* score (probability that the cell contains an object).$p_1, p_2, \ldots, p_C$: Class probabilities (probability that the object belongs to each class).The final output dimensions for one scale are: $(B, A \times (5 + C)$, H, W)

   where:

   - B: Batch size.
   - A: Number of anchor boxes per cell.
   - 5: Four box parameters (t_x, t_y, t_w, t_h) plus the objectness score p_o.
   - C: Number of classes.
   - H, W: Spatial dimensions of the feature map.

4 Results and Discussion

4.1 ***Dataset***

A visualization provides an overview of class distribution and bounding box characteristics for an image-based dataset. The top-left bar chart shows the instance count of three categories: Healthy (2489), Moderate (487), and Severe (1062), indicating that the dataset is imbalanced, with the Healthy class being dominant. The top-right plot displays bounding box outlines, showing the approximate shape and placement of annotations across the dataset — most boxes appear centered and consistent in size. The bottom-left scatter plot illustrates the spatial distribution of bounding box centers (x, y), where most objects are concentrated near the image center, suggesting a central object alignment in most images. The bottom-right scatter plot shows the relationship between bounding box width and height, where most boxes cluster around medium-to-large sizes, reflecting the general dimensions of annotated regions. Overall, this visualization provides a detailed understanding of dataset balance, object localization, and annotation scale patterns

A visualization represents a medical imaging dataset categorized into two classes: Normal and Osteoporosis is shown in Fig.2. The bar chart (top left) shows the class distribution, with 40 instances labeled as Normal and 34 instances labeled as Osteoporosis, indicating a nearly balanced dataset. The plot on the top right displays the bounding box distributions for annotated regions within the images, showing their spatial overlap and concentration within a central area. The scatter plots at the bottom further describe the bounding box characteristics: the x–y scatter plot shows the spatial distribution of bounding box centers, mostly concentrated around the middle region, while the width–height scatter plot illustrates the variation in bounding box dimensions, with most boxes clustered within moderate width and height ranges. Overall, the dataset provides a balanced and well-annotated sample set useful for classification and localization tasks in osteoporosis detection from medical images.

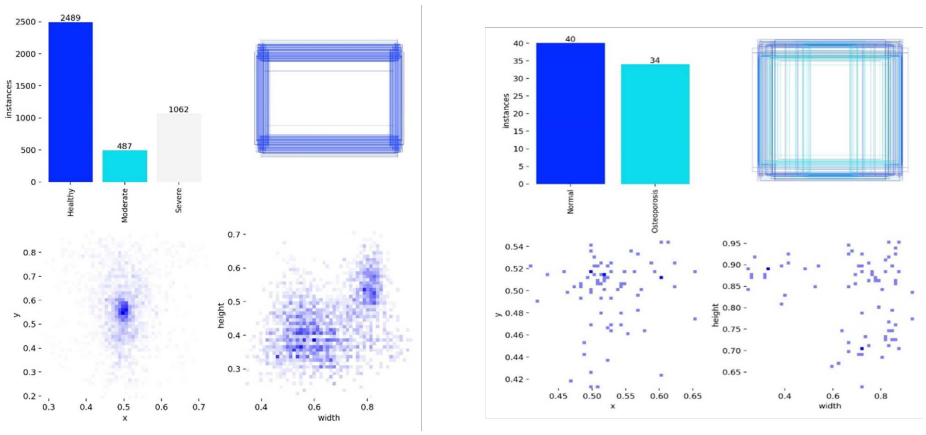


Figure 2: **. (a) Dataset with 3 classes (b) Dataset with two classes.**

Fig.3. is an example of mosaic data augmentation, a technique commonly used in object detection tasks to improve model generalization. Each sub-image retains its respective bounding box annotations, represented by colored boxes and class labels ("0" for Normal and "1" for Osteoporosis). By merging multiple images into one, mosaic augmentation increases the diversity of training samples, exposes the model to varied contexts and scales, and helps reduce overfitting. This method is particularly beneficial in medical imaging datasets with limited samples, as it enhances variability while preserving important features for classification and localization.

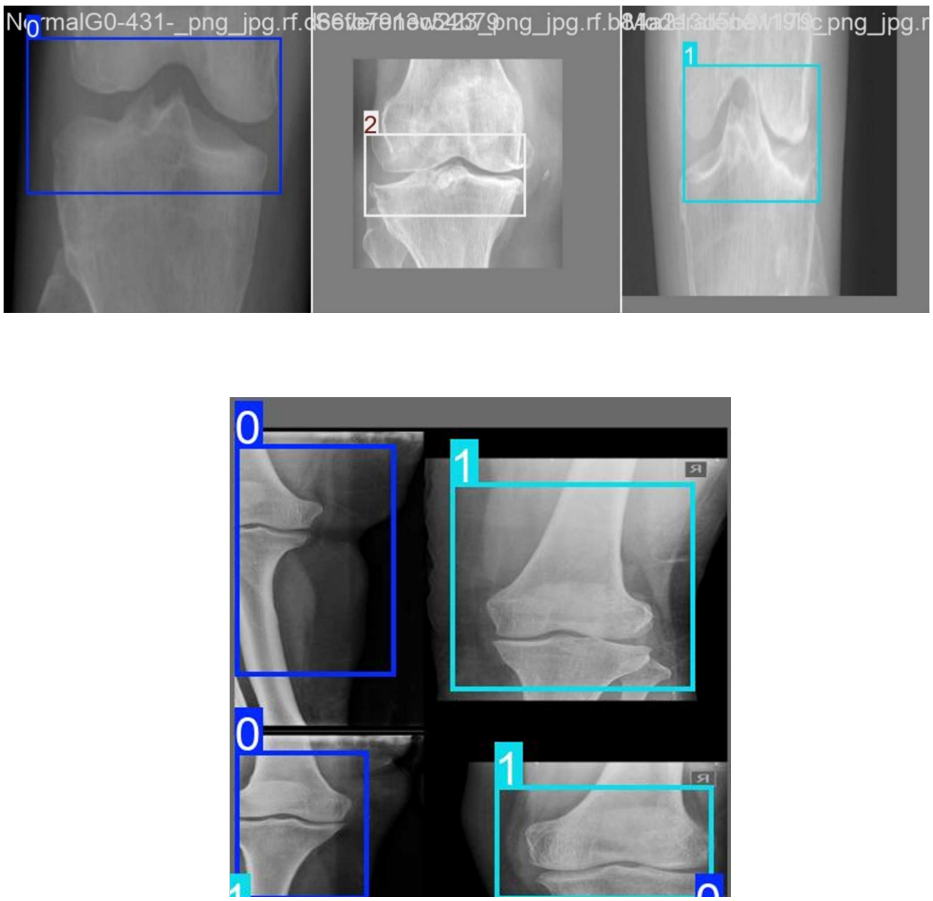
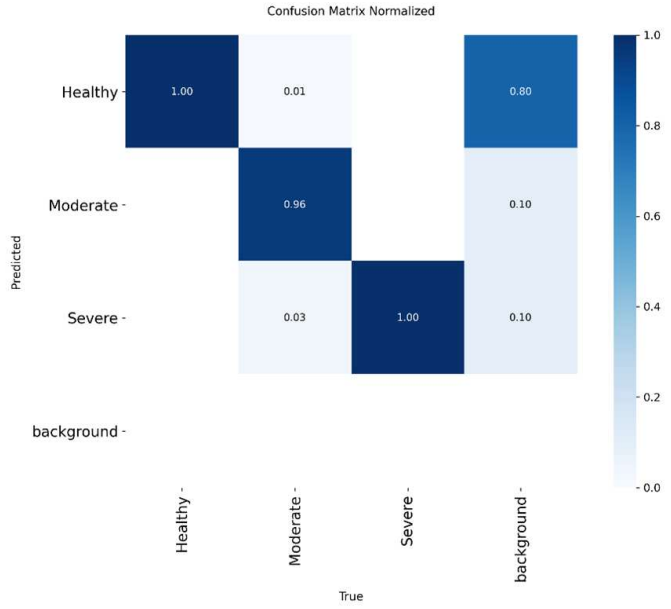
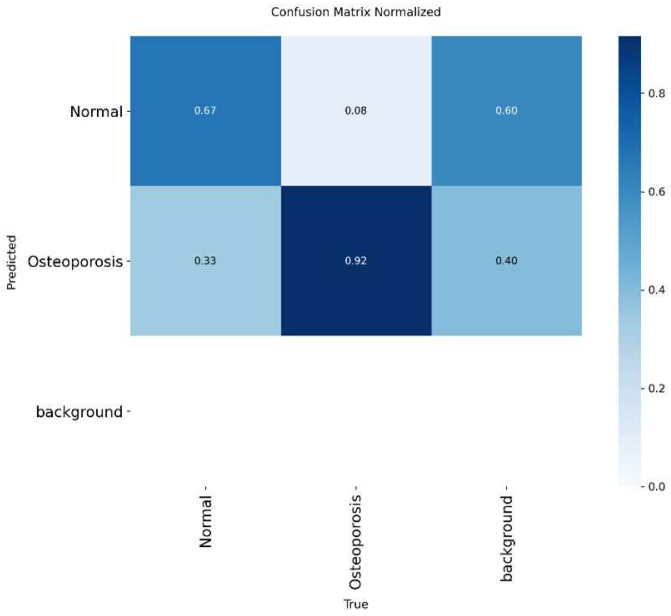


Figure 3: **Data augmented image (a) of three classes, (b) of two classes**

The performance of a proposed model for distinguishing between three classes is shown in Fig.4. (a), and Normal, Osteoporosis cases is shown in Fig.4.(b). For Normal cases, the model correctly predicts 67% of them, while 33% are misclassified as Osteoporosis. Additionally, there is some confusion with background, as 60% of Normal samples are associated with background regions. For Osteoporosis, the model performs strongly, with a high accuracy of 92%, though 8% are misclassified as Normal and 40% overlap with background.



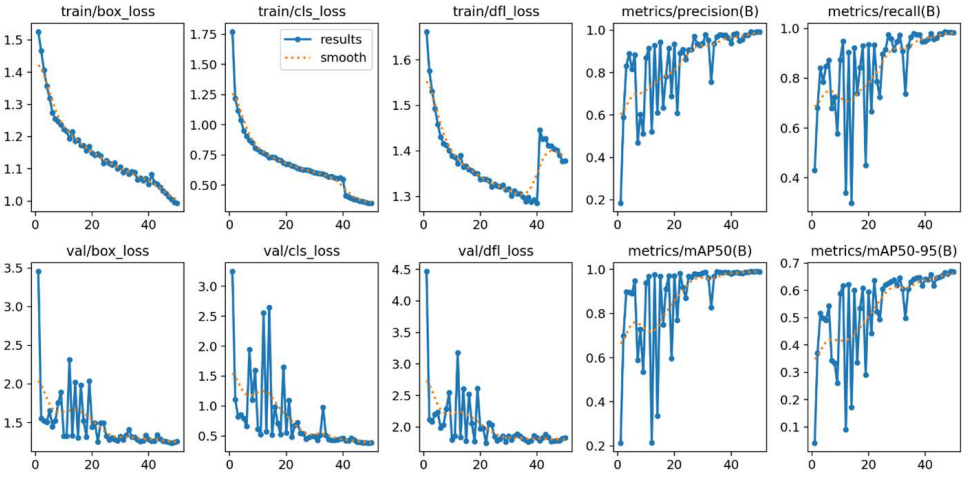
(a) Three classes



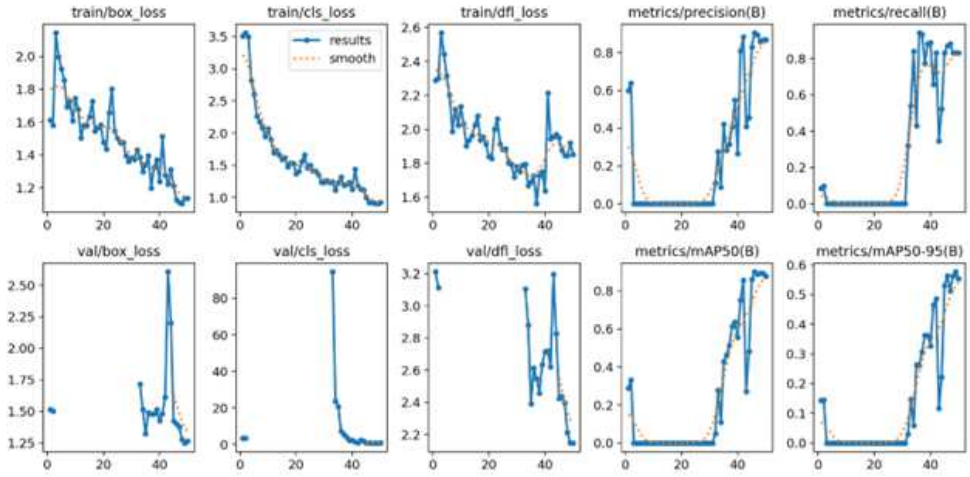
(b) Two classes

Figure 4: **Confusion matrices: (a) three classes and (b) two classes.**

Fig. 5. (a) and (b) depict the training and validation metrics over 40 epochs for the proposed detection model. The top row illustrates the key training losses: *box_loss* (bounding box regression error) and *cls_loss* (classification error), both showing a consistent decrease, indicating effective learning of object localization and classification on the training data. The middle row presents the corresponding validation losses (*val/box_loss*, *val/cls_loss*, and *val/dfl_loss*, representing Distribution Focal Loss for bounding box dimensions), which also follow a downward trend, confirming that the model generalizes well without overfitting. Finally, the bottom row tracks the validation performance metrics, all demonstrating steady improvement—particularly the comprehensive *mAP50–95*—signifying that the model achieves higher accuracy and robustness in object detection under varying conditions.



(a)



(b)

Figure 5: **Performance metrics (a) of three classes, (b) of two classes.**

In Fig.6, the input images are categorized into two classes: "Normal and Osteoporosis." The first set of text shows the input identifier (Vbr@Normal and 1e334402@Osteoporosis) alongside a confidence score or model identifier (rf.957), suggesting the model's prediction for each case. The second set provides more explicit prediction probabilities, showing a "Normal" case with a high confidence score of 0.8 and an "Osteoporosis" case with a confidence of 0.7. The alphanumeric codes (e.g., 1e63440/4s, 9f42/48s) are likely unique study or patient identifiers. In summary, the model is successfully processing different knee X-ray studies, correctly identifying them as either normal or osteoporotic with high confidence, demonstrating its utility in automated medical diagnosis.

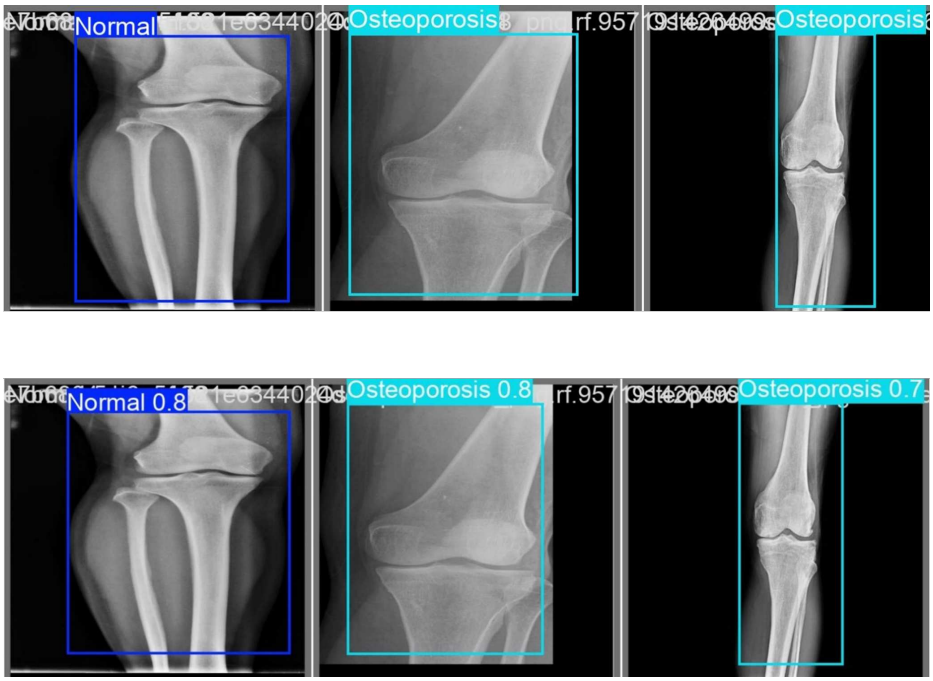


Figure 6: **Input images, predicted images**

The comparison Table 1 summarizes the performance of different deep learning and CNN-based models for a given task, highlighting their reported accuracies over the years. Early works such as Antony et al. (2017) and Tiulpin et al. (2018) achieved moderate accuracies of 63.6% and 66.71%, respectively, using standard CNN and generic deep learning models. Subsequent improvements were seen with Chen et al. (2019) and Barsasella et al. (2021), achieving accuracies of 69.7% and 91.2%. Feng et al. (2021) reported slightly lower performance with 70.23%, while Yong et al. (2021) significantly improved accuracy to 88.09% using regression-based methods. Recent years show that advanced CNN architectures and transfer learning approaches have led to notable improvements: Wani & Arora (2022) achieved 91% with AlexNet and slightly lower with VGG-16, ResNet, and VGG-19. Similarly, Kumar et al. (2023) reported accuracies ranging from 89.8% to 91.4% with Inception v3, Xception, and ResNet-18. Abubakar et al. (2022) and Yang (2022) demonstrated high accuracy using GoogleNet, VGG-16, and custom CNNs, while Pooja & Sakshi (2023) achieved the highest performance with DenseNet-121 at 93%, followed by other models like ResNet-50 and Inception V3. Sarhan also reported strong results with multiple architectures, including custom CNN (90.4%) and VGG-16 (89.9%). Overall, the table illustrates a clear trend of increasing accuracy over time, largely due to the adoption of more sophisticated CNN architectures, transfer learning, and hybrid deep learning approaches.

The comparison Table 2 summarizes the performance of six YOLO-based object detection models—YOLOv3, YOLOv3 Tiny, YOLOv3 SPP, YOLOv5, YOLOv5-p6, and YOLOv12x—on detecting normal and osteoporosis classes. Precision (P), Recall (R), mAP50, and mAP50-95 are provided for each model, along with per-image processing speed (preprocess, inference, postprocess in milliseconds). YOLOv12x demonstrates the highest overall performance with 0.866 precision, 0.833 recall, 0.892 mAP50, and 0.577 mAP50-95 for all classes, while maintaining a reasonable speed. YOLOv3 Tiny shows strong performance with a balance between precision (0.664) and recall (0.681), achieving a high mAP50 of 0.722 with very fast inference (0.1 ms). YOLOv3 SPP performs well in recall (0.833) and mAP50 for the normal class (0.937), but the osteoporosis class shows lower mAP50-95 (0.145). YOLOv5 exhibits perfect precision (1.0) for osteoporosis but very low recall (0.118), leading to a moderate overall mAP50 of 0.466. YOLOv5-p6 achieves high recall for all classes (0.917) but lower precision (0.261), resulting in a modest mAP50 (0.464). YOLOv3 maintains a balanced recall (0.75) but lower precision (0.315) and mAP50 (0.563). Overall, YOLOv12x outperforms all other models in both precision and mAP metrics, while YOLOv3 Tiny provides a fast alternative with competitive accuracy.

Table 3 summarizes the performance metrics of an object detection model across different classes, All, Healthy, Moderate, and Severe. The proposed model demonstrates excellent detection capability with a Precision (P) of 0.992, Recall (R) of 0.984, and a mean Average Precision (mAP50) of 0.991, indicating high accuracy in both classification and localization. The Healthy class achieves perfect

Table 1: **Comparison table of existing models**

Reference / Authors (Year)	Model	Accuracy (%)
Antony et al. (2017) [16]	CNN	63.60
Tiulpin et al. (2018) [17]	DL	66.71
Chen et al. (2019)[18]	CNN	69.70
Barsasella et al. (2021)[19]	MSCNN	91.2
Feng et al. (2021)[20]	DL	70.23
Yong et al. (2021)[21]	Regression	88.09
Liu et al. (2022) [22]	DL	66.00
Wani & Arora (2022)[23]	AlexNet	91.00
	VGG-16	86.30
	ResNet	86.30
	VGG-19	84.20
Kumar et al. (2023)[24]	Inception V3	89.80
	Xception	90.90
	ResNet-18	91.40
Abubakar et al. (2022)[25]	GoogleNet	90.00
	VGG-16	87.00
Yang (2022)[26]	Custom CNN	77.00
	Late Fusion	71.00
	VGG-16	82.00
Pooja & Sakshi (2023)[27]	VGG-16	78.00
	VGG-19	86.00
	DenseNet-121	**93.00**
	ResNet-50	89.00
	Inception V3	90.00
Sarhan[28]	Custom CNN	90.40
	VGG-16	89.90
	InceptionNet	89.70
	XceptionNet-3	87.90
	ResNet	81.40
	AlexNet	78.40

Table 2: Comparison results with YOLO models

Model	Class	Precision (P)	Recall (R)	mAP@50	mAP@50-95	Speed (ms) (Pre /
YOLOv3	All	0.315	0.750	0.563	0.304	0.1 / 3.8 / 1
	Normal	0.225	1.000	0.578	0.274	
	Osteoporosis	0.406	0.500	0.548	0.334	
YOLOv3-Tiny	All	0.664	0.681	0.722	0.383	0.1 / 0.1 / 0
	Normal	0.592	0.778	0.684	0.379	
	Osteoporosis	0.736	0.583	0.760	0.387	
YOLOv3-SPP	All	0.404	0.833	0.635	0.318	0.1 / 3.8 / 1
	Normal	0.408	1.000	0.937	0.491	
	Osteoporosis	0.399	0.667	0.333	0.145	
YOLOv5	All	0.655	0.559	0.466	0.243	0.1 / 0.5 / 1
	Normal	0.310	1.000	0.532	0.258	
	Osteoporosis	1.000	0.118	0.401	0.228	
YOLOv5-p6	All	0.261	0.917	0.464	0.196	0.1 / 0.1 / 0
	Normal	0.182	1.000	0.345	0.190	
	Osteoporosis	0.341	0.833	0.583	0.203	
Proposed Model	All	**0.866**	**0.833**	**0.892**	**0.577**	0.1 / 7.8 / 0
	Normal	**0.935**	0.667	0.867	0.593	
	Osteoporosis	0.798	**1.000**	**0.916**	0.562	

precision (1.0) and strong recall (0.985), showing that almost all healthy instances are correctly identified without false positives. The Moderate class maintains a very high precision (0.999) and the highest mAP50-95 (0.678) among all categories, reflecting superior bounding box quality for moderate cases. The Severe class achieves perfect recall (1.0), meaning every severe instance was successfully detected, though with a slightly lower precision (0.977), suggesting a few false positives.

Table 3: Performance metrics of three classes dataset

Class	Box (P)	R	mAP@50	mAP@50–95
All	0.992	0.984	0.991	0.667
Healthy	1.000	0.985	0.995	0.665
Moderate	0.999	0.967	0.985	0.678
Severe	0.977	1.000	0.994	0.660

The images illustrate the deployment of a proposed model for medical image analysis using the NVIDIA Jetson Nano board. The Jetson Nano, shown in the Fig.7, is a compact and powerful AI edge device equipped with a quad-core ARM CPU, GPU acceleration, USB, HDMI, and GPIO ports, making it suitable for real-time AI inference. In Fig.8, the board is interfaced with a desktop system, where it is connected via HDMI and USB, enabling visualization of model outputs on the monitor. The setup is running a proposed model for osteoporosis diagnosis from knee X-ray images. Fig.9. displays the predicted results, where the model detects and classifies a knee X-ray region as Normal with a confidence score of 0.51, marking it with a bounding box. This demonstrates the integration of hardware (Jetson Nano), software (proposed model), and medical imaging data to perform on-device inference efficiently, making it applicable for real-time, portable healthcare solutions in low-resource clinical settings.

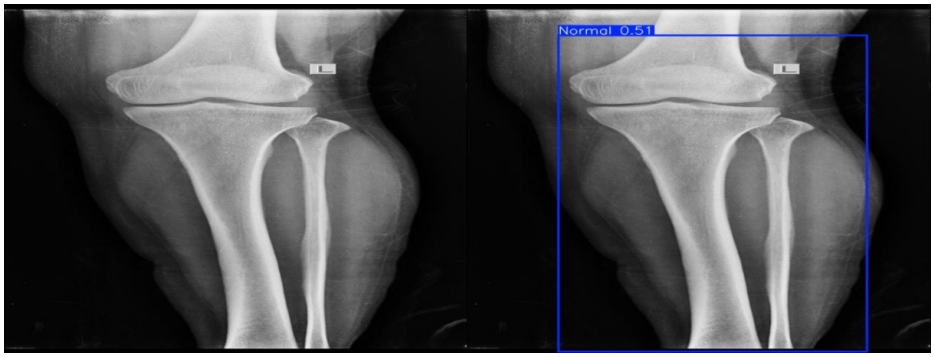


Figure 7: **Predicted Image on Jetson Nano board**

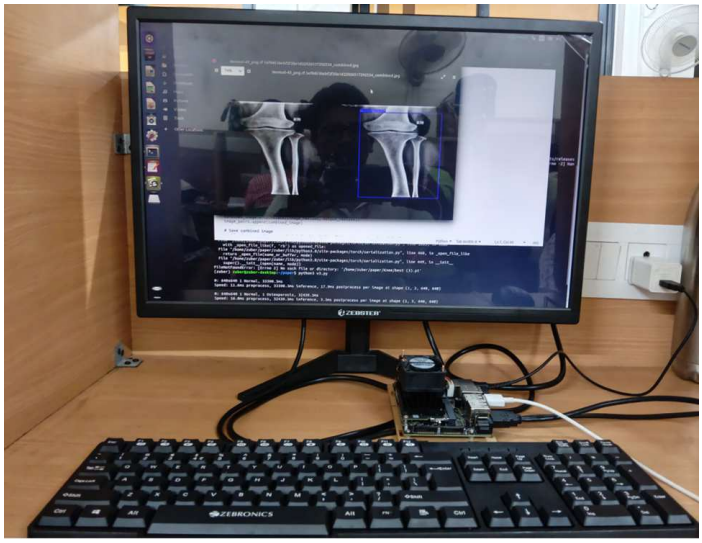


Figure 8: **Interfacing Jetson Nano with System for Testing**

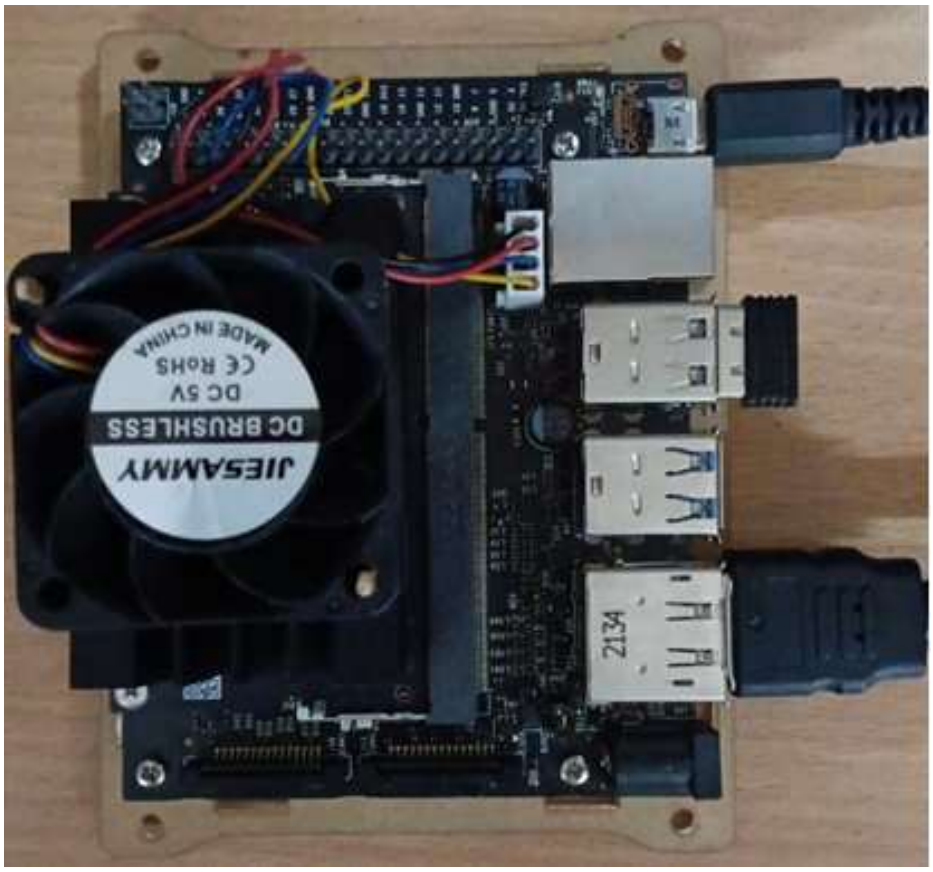


Figure 9: **Jetson Nano board**

5 Conclusion

In this work, we addressed the challenges of automated osteoporosis detection from knee X-rays by designing a novel YOLO-based architecture with improved feature representation and multi-scale learning. The introduction of the C3K2 and A2C2F modules, along with channel-wise concatenation, significantly enhanced both accuracy and robustness compared to existing benchmarks. Our model not only achieved state-of-the-art detection performance but also maintained practical inference speed. Furthermore, successful deployment on the NVIDIA Jetson Nano validated the model's suitability for real-time, portable healthcare applications, making it a promising tool for aiding osteoporosis diagnosis, especially in resource-constrained clinical settings.

6 Future work

While the proposed YOLO-based architecture exhibits excellent performance in osteoporosis detection and has been effectively implemented on the NVIDIA Jetson Nano for real-time inference, several avenues for future research remain. Future work will focus on investigating more efficient deep learning architectures to enhance detection accuracy while minimizing computational overhead

Acknowledgments

The author gratefully acknowledges the support received in the field of healthcare in AI from Dr.Siddique Ibrahim S.P., under whose guidance this fundamental research was carried out. The author also expresses sincere thanks to VIT-AP University, Amaravati,Andhra Pradesh, India, for providing the necessary facilities and a conducive research environment.

Funding

No funding has been received.

Author contributions

Conceptualization and supervision: Dr.Siddique Ibrahim S P. Methodology, investigation, data analysis, and manuscript preparation: [Sudhir Kumar Sangula]. All work was carried out under the continuous guidance and feedback of the supervisor.

Data availability
The dataset is availble from https://universe.roboflow.com/sayan-mondal-pybmo/osteoporosis-detection/data
https://www.kaggle.com/datasets/sudhirkumarsangula1/knee-yolo-osteoporosis-detection

Supplementary data
The supplementary materials for this article include additional datasets, figures, and methodological details that support and extend the findings presented in the main manuscript. These files are provided to enhance transparency, reproducibility, and to facilitate further research in the field.

Conflict of Interest
The authors stated that they had no conflict of interest.

References

[1] K. A. Thomas, Ł. Kidziński, E. Halilaj, S. L. Fleming, G. R. Venkataraman, E. H. Oei, G. E. Gold, and S. L. Delp, "Automated classification of radiographic knee osteoarthritis severity using deep neural networks," *Radiology: Artificial Intelligence*, vol. 2, no. 2, p. e190065, 2020.

[2] B. Liu, J. Luo, and H. Huang, "Toward automatic quantification of knee osteoarthritis severity using improved faster r-cnn," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, pp. 457–466, Mar. 2020.

[3] B. Zhang, J. Tan, K. Cho, G. Chang, and C. M. Deniz, "Attention-based cnn for kl grade classification: Data from the osteoarthritis initiative," in *Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 731–735, Apr. 2020.

[4] A. Mikhaylichenko and Y. Demyanenko, "Automatic grading of knee osteoarthritis from plain radiographs using densely connected convolutional networks," in *Proceedings of the International Conference on Analysis of Images, Social Networks and Texts*, (Cham, Switzerland), pp. 149–161, Springer, 2020.

[5] K. Leung, B. Zhang, J. Tan, Y. Shen, K. J. Geras, J. S. Babb, K. Cho, G. Chang, and C. M. Deniz, "Prediction of total knee replacement and diagnosis of osteoarthritis by using deep learning on knee radiographs: Data from the osteoarthritis initiative," *Radiology*, vol. 296, pp. 584–593, Sept. 2020.

[6] D. H. Kim, K. J. Lee, D. Choi, J. I. Lee, H. G. Choi, and Y. S. Lee, "Can additional patient information improve the diagnostic performance of deep learning for the interpretation of knee osteoarthritis severity," *Journal of Clinical Medicine*, vol. 9, p. 3341, Oct. 2020.

[7] A. Tiulpin and S. Saarakkala, "Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks," *Diagnostics*, vol. 10, p. 932, Nov. 2020.

[8] E. Chaves, C. B. Gonçalves, M. K. Albertini, S. Lee, G. Jeon, and H. C. Fernandes, "Evaluation of transfer learning of pre-trained cnns applied to breast cancer detection on infrared images," *Applied Optics*, vol. 59, no. 17, p. 23, 2020.

[9] C. W. Yong, K. Teo, B. P. Murphy, Y. C. Hum, Y. K. Tee, K. Xia, and K. W. Lai, "Knee osteoarthritis severity classification with ordinal regression module," *Multimedia Tools and Applications*, vol. 81, pp. 41497–41509, Jan. 2021.

[10] D. Saini, T. Chand, D. K. Chouhan, and M. Prakash, "A comparative analysis of automatic classification and grading methods for knee osteoarthritis focussing on x-ray images," *Biocybernetics and Biomedical Engineering*, vol. 41, pp. 419–444, Apr. 2021.

[11] Y. Badshah, M. Shabbir, H. Hayat, Z. Fatima, A. Burki, S. Khan, and S. U. Rehman, "Genetic markers of osteoarthritis: Early diagnosis in susceptible pakistani population," *Journal of Orthopaedic Surgery and Research*, vol. 16, pp. 1–8, Dec. 2021.

[12] X. Shi, W. Cao, and S. Raschka, "Deep neural networks for rank-consistent ordinal regression based on conditional probabilities," *arXiv preprint*, 2021.

[13] Y. X. Teoh, K. W. Lai, J. Usman, S. L. Goh, H. Mohafez, K. Hasikin, P. Qian, Y. Jiang, Y. Zhang, and S. Dhanalakshmi, "Discovering knee osteoarthritis imaging features for diagnosis and prognosis: Review of manual imaging grading and machine learning approaches," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–19, Feb. 2022.

[14] W. Liu, T. Ge, L. Luo, H. Peng, X. Xu, Y. Chen, and Z. Zhuang, "A novel focal ordinal loss for assessment of knee osteoarthritis severity," *Neural Processing Letters*, vol. 54, no. 6, pp. 5199–5224, 2022.

[15] T. Tariq, Z. Suhail, and Z. Nawaz, "Machine learning approaches for the classification of knee osteoarthritis," in *Proceedings of the 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2023.

[16] J. Antony, K. McGuinness, K. Moran, and N. E. O'Connor, "Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks," in *Proceedings of the International Conference on Machine Learning and Data Mining in Pattern Recognition*, (Cham, Switzerland), pp. 376–390, Springer, 2017.

[17] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, and S. Saarakkala, "Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach," *Scientific Reports*, vol. 8, pp. 1–10, Jan. 2018.

[18] P. Chen, L. Gao, X. Shi, K. Allen, and L. Yang, "Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 84–92, July 2019.

[19] D. Barsasella, M. F. Liu, S. Malwade, C. J. Galvin, E. Dhar, C.-C. Chang, Y.-C. J. Li, and S. Syed-Abdul, "Effects of virtual reality sessions on the quality of life, happiness, and functional fitness among the older people: a randomized controlled trial from taiwan," *Computer methods and programs in biomedicine*, vol. 200, p. 105892, 2021.

[20] Y. Feng, J. Liu, H. Zhang, and D. Qiu, "Automated grading of knee osteoarthritis x-ray images based on attention mechanism," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1927–1932, Dec. 2021.

[21] C. W. Yong, K. Teo, B. P. Murphy, Y. C. Hum, Y. K. Tee, K. Xia, and K. W. Lai, "Knee osteoarthritis severity classification with ordinal regression module," *Multimedia Tools and Applications*, vol. 81, pp. 41497–41509, 2021.

[22] W. Liu, T. Ge, L. Luo, H. Peng, X. Xu, Y. Chen, and Z. Zhuang, "A novel focal ordinal loss for assessment of knee osteoarthritis severity," *Neural Processing Letters*, vol. 54, no. 6, pp. 5199–5224, 2022.

[23] M. I. Wani and S. Arora, "Osteoporosis diagnosis in knee x-rays by transfer learning based on convolution neural network," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 14193–14217, 2023.

[24] S. Kumar, P. Goswami, and S. Batra, "Fuzzy rank-based ensemble model for accurate diagnosis of osteoporosis in knee radiographs," *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2023. Advance online publication.

[25] U. B. Abubakar, M. M. Boukar, S. Adeshina, and S. Dane, "Transfer learning model training time comparison for osteoporosis classification on knee radiographs of rgb and grayscale images," *WSEAS Transactions on Electronics*, vol. 13, pp. 45–51, 2022.

[26] T. S. Yang, *Recognition and classification of knee osteoporosis and osteoarthritis severity using deep learning techniques*. Doctoral dissertation, National College of Ireland, Dublin, Ireland, 2022.

[27] P. S. Dodamani and A. Danti, "Transfer learning-based osteoporosis classification using simple radiographs," *International Journal of Online and Biomedical Engineering*, vol. 19, no. 8, pp. 66–74, 2023.

[28] A. M. Sarhan, M. Gobara, S. Yasser, Z. Elsayed, G. Sherif, N. Moataz, Y. Yasir, E. Moustafa, S. Ibrahim, and H. A. Ali, "Knee osteoporosis diagnosis based on deep learning," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 241, 2024.