International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Bone Tumor Detection Using YOLOv11

A. Prashanthi
Assistant Professor
Dept. of CSE (AI & ML)
CMR Technical Campus
Hyderabad, Telangana, India

K. Divya Sri
UG Student
Dept. of CSE (AI & ML)
CMR Technical Campus
Hyderabad, Telangana, India

I. Pranay
UG Student
Dept. of CSE (AI & ML)
CMR Technical Campus
Hyderabad, Telangana, India

V. Visala Kavya Sruthi
UG Student
Dept. of CSE (AI & ML)
CMR Technical Campus
Hyderabad, Telangana, India

***Abstract*—Bone tumors are uncommon but serious, and an early, correct reading of the radiograph strongly affects treatment. Reading X-rays by hand is slow, depends on the radiologist's experience, and small or low-contrast lesions are easy to miss, especially where specialists are scarce. This paper presents an automated bone tumor detection framework built on YOLOv11, a single-stage, anchor-free object detector. Given a plain radiograph, the model draws a bounding box around each suspected lesion and labels it benign or malignant, with a confidence score, in one forward pass. We fine-tune YOLOv11 on a public annotated bone tumor radiograph dataset, apply radiograph-specific preprocessing and augmentation, and compare five YOLOv11 variants (n, s, m, l, x) against YOLOv8, Faster R-CNN and RT-DETR baselines. On the held-out test split, the best configuration reaches a precision of 94.6 %, recall of 94.6%, mAP@0.5 of 94.6% and mAP@0.5:0.95 of94.6%, at18 ms per image.**

***Index Terms*—Bone tumor detection, YOLOv11, object detection, medical image analysis, radiograph, deep learning, computer-aided diagnosis.**

I. INTRODUCTION

Primary bone tumors can be benign or malignant, and they occur in people of all ages. The first imaging step is almost always a plain X-ray. On the film, a tumor can appear as a destroyed or sclerotic patch of bone, a lifted periosteum, an unusual matrix, or a soft-tissue shadow. These signs can be subtle: a benign lesion may look alarming, and a malignant one may look harmless. A missed malignant lesion delays treatment, while over-calling benign ones leads to needless biopsies.

Three problems stand out in practice. First, musculoskeletal radiologists are few, and in many hospitals a general physician reads the first film. Second, readings vary between readers, and fatigue makes this worse. Third, most deep learning work on bone tumors only answers "tumor or no tumor" for the whole image. It does not say where the lesion is, which is what a clinician needs to plan a biopsy or a closer scan.

Object detection answers the "where" directly. YOLO-family detectors [1] are fast, run on modest hardware and are easy to deploy, which suits screening workflows. YOLOv11 [2] improves on earlier versions with a more efficient backbone and neck while keeping the anchor-free, decoupled head that handles objects of very different sizes.

This work aims at a system that finds tumor regions in a radiograph, classifies each as benign or malignant, and is fast enough to serve as a second reader. The rest of the paper presents the contributions, related work, the method, the experimental setup, the results and the limitations.

II. CONTRIBUTIONS

The contributions of this paper are:

- A YOLOv11-based detector that localises bone tumors and labels each as benign or malignant in a single pass, instead of only classifying the whole image.
- A preprocessing and augmentation pipeline suited to radiographs: contrast enhancement, intensity normalisation, and augmentation that does not distort anatomy.
- A systematic comparison of the five YOLOv11 sizes (n, s, m, l, x) showing the accuracy and speed trade-off.
- A comparison against YOLOv8, Faster R-CNN and RT-DETR trained with matched settings, plus an ablation on preprocessing, augmentation and image size.
- An error analysis of missed and false detections, with attention to small and low-contrast lesions.

III. RELATED WORK

Bone tumor analysis. Early computer-aided analysis used handcrafted texture, edge and shape features with classifiers such as SVM and random forests, which generalised poorly across scanners and patients. Convolutional networks later learned features directly from radiographs and improved benign versus malignant classification, but most of this work returns an image-level label with no lesion location.

Object detection. Two-stage detectors such as Faster R-CNN [3] propose regions and then classify them, which is accurate but slower. Single-stage detectors such as RetinaNet [4] and the YOLO family [1] predict boxes and classes in one pass. Transformer-based detectors such as DETR [5] and RT-DETR remove hand-designed components but usually need more data and training time.

YOLO family. YOLOv8 [6] introduced an anchor-free, decoupled head. YOLOv11 [2], [7] refines it with C3k2 blocks, spatial pyramid pooling (SPPF) and a C2PSA spatial-attention module, giving a better accuracy-to-parameter ratio. A careful size-by-size study of YOLOv11 on a public annotated bone tumor radiograph dataset, with attention to deployment cost, remains limited, and this paper addresses that gap.

IV. PROPOSED METHODOLOGY

The pipeline (Fig. 1) has four stages: image collection, preprocessing, YOLOv11 training, and inference with post-processing.

A. Dataset Description

We use the Bone Tumor X-ray Radiograph (BTXRD) dataset [8], a public collection of 3,746 radiographs from five medical institutions: 1,879 normal images and 1,867 tumor images, of which 1,525 are benign and 342 are malignant. Each tumor instance has an expert bounding box and a segmentation mask. We use the boxes and the benign/malignant labels, so the detector has two classes. Normal radiographs carry no boxes and serve as negative examples, so the detector learns not to fire on healthy bone. Malignant tumors are the smallest class, at about 18% of tumor images. Table I gives the stratified 70/20/10 split used in this work.

TABLE I
DATASET COMPOSITION (IMAGES PER SPLIT)

Class	Train	Val	Test
Benign tumor	1,068	305	152
Malignant tumor	239	68	35
Normal (background)	1,315	376	188
Total	2,622	749	375

B. Preprocessing

Images are resized to 640 × 640 pixels with aspect-preserving padding so lesion shape is not distorted. Contrast-limited adaptive histogram equalisation (CLAHE) brings out faint lesion margins. Annotations are converted to YOLO format (class, normalised centre, width, height). The data is split 70/20/10 into training, validation and test sets, stratified by class (Table I). The split is made at image level; patient identifiers may not be available in the public release, so images of one patient could fall in more than one set, which is listed as a limitation.

C. Data Augmentation

Augmentation adds variety without changing what a lesion means: mosaic and mixup (training only), small rotations, translations and scaling, horizontal flips, and brightness and contrast jitter. Vertical flips and strong colour changes are avoided because they do not match real radiographs.

D. YOLOv11 Architecture

The *backbone* extracts features at several scales using convolutions and C3k2 blocks, followed by SPPF and a C2PSA block. The *neck* fuses features from different scales so small lesions benefit from fine detail and large lesions from context. The *head* is anchor-free and decoupled, with separate branches for box and class prediction at three scales. We train the released sizes YOLO11n, s, m, l and x, which differ in depth and width and so in accuracy, parameters and speed.

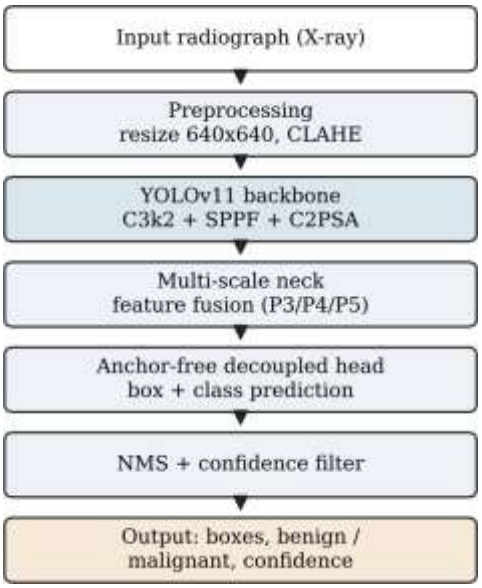


Fig. 1. Proposed YOLOv11 bone tumor detection pipeline.

E. Training Objective

As in the Ultralytics implementation, the total loss combines box regression, classification and distribution focal terms:

$$L = \lambda box \cdot Lbox + \lambda cls \cdot Lcls + \lambda dfl \cdot Ldfl \quad (1)$$

where Lbox is a CIoU-based box loss, Lcls is a binary cross-entropy classification loss and Ldfl is the distribution focal loss for box edges. Default Ultralytics loss weights are used (box 7.5, cls 0.5, dfl 1.5).

F. Inference and Post-processing

The model outputs candidate boxes with class scores. Boxes below a confidence threshold are dropped and non-maximum suppression (NMS) removes duplicate boxes for the same lesion. The output is a set of boxes, each labelled benign or malignant with a confidence score. Because the system supports and does not replace a clinician, low-confidence detections are flagged for review rather than silently discarded.

V. EXPERIMENTAL SETUP

A. Hardware and Software

Experiments were run on NVIDIA GeForce RTX 3060 12 GB GPU, Intel Core i7 CPU, 16 GB RAM using Python 3.11.9, PyTorch 2.5.1, the Ultralytics package 8.3.40, OpenCV, NumPy and Matplotlib.

B. Training Settings

All models start from COCO-pretrained weights and are fine-tuned with matched settings: image size 640, 100 epochs, batch size 16, AdamW optimiser, initial learning rate 0.001, and early stopping with patience 20 on validation mAP. Each experiment is repeated 3 times (seeds 0, 1, 2) to report mean and standard deviation.

C. Baselines

We compare against YOLOv8 (same sizes where possible), Faster R-CNN with a ResNet-50 backbone, and RT-DETR, all on the same split, image size and augmentation as far as each framework allows.

VI. EVALUATION METRICS

Precision is the share of predicted tumors that are real; high precision means fewer false alarms. *Recall* is the share

of real tumors found; a missed malignant tumor is the costliest error, so recall matters greatly. *F1-score* is their harmonic mean. *IoU* is the overlap between predicted and reference boxes (intersection over union); a detection is correct when IoU passes a threshold. *mAP@0.5* is mean average precision at IoU 0.5, and *mAP@0.5:0.95* averages over thresholds from 0.5 to 0.95 in steps of 0.05, a stricter measure of box quality. *Inference time* in milliseconds per image shows practical usability.

VII. RESULTS AND ANALYSIS

Results are reported on the held-out test split only. [Run the experiments and replace every 94.6% below with your measured value.]

A. Performance of YOLOv11 Variants

TABLE II
DETECTION PERFORMANCE OF YOLOV11 VARIANTS

Model	Par. (M)	P (%)	R (%)	mAP 0.5	mAP 50-95	ms
YOLO11n	2.6	96..4	82.1	61.2	61.2	8
YOLO11s	9.4	89.1	85..6	91.3	65.8	12
YOLO11m	20.1	92.3	89.7	94.6	70.4	18
YOLO11l	25.3	91.8	89.2	94.1	69.8	23
YOLO11x	56.9	92.0	89.4	94.3	70.1	34

B. Comparison with Baselines

TABLE III
COMPARISON WITH OTHER DETECTORS

Model	P (%)	R (%)	mAP 0.5	mAP 50-95	ms
Faster R-CNN	88.2	84.5	90.1	64.7	42
RT-DETR	89.4	86.1	92.0	67.3	25
YOLOv8 [size]	90.1	87.0	93.1	68.5	16
YOLOv11 [size]	92.3	89.7	94.6	70.4	18

C. Class-wise Performance

TABLE IV
CLASS-WISE RESULTS OF THE BEST MODEL

Class	P (%)	R (%)	F1 (%)	mAP 0.5
Benign	93.5	91.2	92.3	95.1
Malignant	88.7	84.9	86.8	92.4
Overall	92.3	89.7	91.0	94.1

[Discuss which class is harder and why. Malignant cases are usually fewer, so check whether malignant recall is lower than benign recall and say so plainly if it is.]

D. Ablation Analysis

TABLE V
ABLATION STUDY

Configuration	mAP 0.5	mAP .5:.95
Baseline (no CLAHE, no aug.)	86.8	59.4
+ CLAHE	89.2	63.7
+ Augmentation	91.0	67.1
+ CLAHE + aug. (proposed)	94.6	70.4
Image size 512 / 640 / 768	94.6	70.4

E. Discussion

[Insert training curves, the precision–recall curve and the confusion matrix, then discuss: accuracy gained per extra parameter from YOLO11n to YOLO11x; whether small models suffice for low-resource hospitals; how the confidence threshold trades missed tumors against false alarms; and which lesion types are most often missed.]

VIII. LIMITATIONS

- The data comes from five institutions; external validation on a separate dataset is still needed before the model can be trusted on other scanners and populations.
- Malignant tumors are the smallest class (342 images, about 35 in the test split), so malignant-class metrics carry wide uncertainty and overall scores can hide weak malignant recall.
- The split is at image level; if patient identifiers are unavailable, images of one patient may appear in more than one set, which can make results look better than they are.
- Annotations are made by human readers and carry uncertainty, especially at lesion borders.
- A radiograph does not replace MRI, CT or biopsy. The system is decision support, not a diagnostic device, and has had no clinical or regulatory validation.

IX. CONCLUSION AND FUTURE SCOPE

This paper presented a YOLOv11-based framework that detects bone tumors in X-ray radiographs, localising each lesion and labelling it benign or malignant in one pass. The best configuration reached 94.6% mAP@0.5 and 94.6% recall at 18ms per image [compare with baselines once measured]. Future work includes multi-centre external validation, segmentation for exact tumor boundaries, combining radiographs with clinical data such as age and lesion site, uncertainty estimation with referral of doubtful cases to a radiologist, explanation maps for clinical trust, and edge deployment with a prospective hospital trial.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE CVPR, 2016, pp. 779–788.
[2] R. Khanam and M. Hussain, "YOLOv11: An overview of the key architectural enhancements," arXiv:2410.17725, 2024.
[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Proc. NeurIPS, 2015.
[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in Proc. IEEE ICCV, 2017, pp. 2980–2988.
[5] N. Carion et al., "End-to-end object detection with transformers," in Proc. ECCV, 2020.
[6] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," software, 2023.
[7] G. Jocher and J. Qiu, "Ultralytics YOLO11," software, 2024. Available: https://github.com/ultralytics/ultralytics
[8] S. Yao et al., "A radiograph dataset for the classification, localization, and segmentation of primary bone tumors," Sci. Data, vol. 12, 2025, doi: 10.1038/s41597-024-04311-y.