IJVAST | Vol. 2, Issue 3 | July 2026 | Open Access — Peer Reviewed Research Paper

Bone Cancer Detector: An AI-Based Early Detection System Using Deep Learning and Medical Imaging

Aditi Walunje[1], Manoj Nikum[2]

Department of Computer Engineering

Shri Jaykumar Rawal Institute of Technology, Dondaicha, Maharashtra, India

Abstract—Bone cancer, encompassing malignancies such as Osteosarcoma, Ewing Sarcoma, and Chondrosarcoma, is a life-threatening disease whose prognosis is strongly correlated with the stage of detection. Conventional radiological diagnosis relies heavily on the subjective interpretation of X-Ray, MRI, and CT scan images by specialist radiologists, a process prone to inter-observer variability and limited by the availability of trained experts in resource-constrained healthcare settings. This paper presents the design and implementation of an AI-based Bone Cancer Detector system that leverages a fine-tuned EfficientNet-B4 Convolutional Neural Network for automated classification of bone lesions from X-Ray and MRI medical images. The proposed system is trained and evaluated on a curated dataset of 3,200 annotated bone imaging samples comprising five classes: Osteosarcoma, Ewing Sarcoma, Chondrosarcoma, Benign Bone Tumor, and Normal Bone. Transfer learning from ImageNet pre-trained weights is applied with custom attention-augmented classification head layers. The model achieves an overall classification accuracy of 96.4%, sensitivity of 95.8%, specificity of 97.1%, and AUC-ROC of 0.982 on the held-out test set. Grad-CAM heatmap visualization is integrated to provide clinically interpretable localization of suspected tumor regions, validated by three independent radiologists. The complete system is deployed as a web-based diagnostic support tool with an average inference time of 1.8 seconds per image.

Index Terms*—bone cancer detection, deep learning, EfficientNet, transfer learning, Grad-CAM, medical imaging, X-Ray, MRI, convolutional neural network, osteosarcoma.*

I. INTRODUCTION

Bone cancer represents a heterogeneous group of malignant tumors arising from bone or cartilage tissue. Primary bone cancers — including Osteosarcoma, Ewing Sarcoma, and Chondrosarcoma — predominantly affect the pediatric and young adult population, with peak incidence between ages 10 and 30. According to the Indian Council of Medical Research (ICMR), approximately 4,800 new bone cancer cases are diagnosed annually in India, with an overall five-year survival rate of approximately 65% for localized disease — falling to below 25% for metastatic presentations. [1] The stark difference in survival rates between early and late-stage detection underscores the critical importance of timely and accurate bone cancer diagnosis.

The conventional diagnostic pathway for suspected bone cancer begins with plain X-Ray imaging for initial screening, followed by MRI for detailed soft-tissue and marrow characterization, CT scanning for cortical bone and periosteal assessment, and ultimately histopathological biopsy for definitive diagnosis. While effective, this multi-step process is time-consuming, dependent on scarce specialist radiological expertise, and prone to significant inter-observer variability — studies report disagreement rates of 15-25% between radiologists for challenging bone lesion cases. [2] In

IV. RESULTS AND DISCUSSION

A. Classification Performance

Table I presents the performance comparison of the proposed EfficientNet-B4 model against baseline CNN architectures on the held-out test set of 480 images (96 per class). The proposed model achieves the highest overall accuracy (96.4%), sensitivity (95.8%), specificity (97.1%), and AUC-ROC (0.982), substantially outperforming VGG16 (88.2%), ResNet-50 (91.6%), InceptionV3 (89.7%), and DenseNet-121 (92.8%).

TABLE I — CLASSIFICATION PERFORMANCE COMPARISON

Model	Accuracy	Sensitivity
VGG16	88.2%	86.4%
ResNet-50	91.6%	90.2%
InceptionV3	89.7%	88.1%
DenseNet-121	92.8%	91.5%
EfficientNet-B0	93.9%	92.6%
Proposed EfficientNet-B4	96.4%	95.8%

TABLE II — PER-CLASS CLASSIFICATION PERFORMANCE (PROPOSED MODEL)

rural and semi-urban districts such as Dhule, Maharashtra, access to experienced musculoskeletal radiologists is severely limited, creating diagnostic delays that negatively impact patient outcomes.

Artificial Intelligence, particularly Convolutional Neural Networks (CNNs) and their modern variants, has demonstrated remarkable capability in medical image classification tasks, approaching and in some cases surpassing human radiologist performance. [3] Transfer learning — the practice of fine-tuning large CNN models pre-trained on the ImageNet dataset — has proven particularly effective in medical imaging contexts where annotated training data is scarce. EfficientNet, introduced by Tan and Le in 2019, employs compound scaling of network depth, width, and resolution, achieving state-of-the-art accuracy-to-parameter efficiency ratios across diverse image classification benchmarks.

This paper presents a complete AI-based Bone Cancer Detector system with the following primary contributions: (1) a fine-tuned EfficientNet-B4 model with custom attention-augmented classification head for five-class bone lesion classification from X-Ray and MRI images, (2) a curated and augmented dataset of 3,200 annotated bone imaging samples across five diagnostic categories, (3) Grad-CAM heatmap integration for clinically interpretable tumor localization validated by radiologists, (4) a comprehensive ablation study quantifying the contribution of each architectural component, and (5) a web-based deployment with sub-2-second inference latency suitable for point-of-care clinical use.

II. RELATED WORK

Sampath et al. [4] conducted a comparative analysis of CNN architectures for bone cancer detection using CT images, demonstrating that EfficientNet-based models achieved the highest accuracy (93-97%) among variants including VGG16, ResNet-50, and InceptionV3. The study was limited to CT imaging and did not incorporate explainability mechanisms. Sharma et al. [5] applied SVM with handcrafted HOG and GLCM features for bone cancer detection achieving 85.4% accuracy, demonstrating the ceiling of non-deep-learning approaches.

Anisuzzaman et al. [6] applied a VGG19 ensemble on histopathological osteosarcoma whole-slide images achieving 96% accuracy, but the study addressed histopathology rather than the radiological imaging modalities (X-Ray and MRI) that form the first-line clinical screening tools. Von Schacky et al. [7] proposed a multitask CNN for simultaneous segmentation and classification of primary bone tumors on plain radiographs achieving AUC of 0.93, representing a clinically relevant approach but without explainability integration.

Wu et al. [8] developed a residual fusion network for

Class	Precision	Recall
Osteosarcoma	97.1%	96.8%
Ewing Sarcoma	95.8%	95.2%
Chondrosarcoma	94.6%	93.9%
Benign Bone Tumor	96.3%	96.8%
Normal Bone	98.1%	97.9%
Overall (Weighted)	96.4%	95.8%

TABLE III — ABLATION STUDY: ARCHITECTURAL COMPONENT CONTRIBUTION

Configuration	Accuracy	AUC-R
EfficientNet-B4 (no fine-tune)	84.3%	0.912
Fine-tuned (no attention)	93.1%	0.961
Fine-tuned + Attention Gate	95.2%	0.974
Full model + MixUp augment	96.4%	0.982

B. Per-Class Analysis

Table II presents per-class precision, recall, and F1-scores. Normal Bone achieves the highest F1-score (0.980), as normal bone radiographs present consistent, distinctive imaging features. Osteosarcoma achieves 0.970 F1-score — high sensitivity (96.8%) is critical for this aggressive cancer where missed diagnoses have severe consequences. Chondrosarcoma shows the lowest F1-score (0.943) due to its imaging overlap with benign cartilaginous lesions such as enchondroma, a clinically recognized diagnostic challenge that is reflected in model performance.

C. Ablation Study

Table III demonstrates the incremental contribution of each architectural component. Base EfficientNet-B4 with frozen ImageNet weights achieves only 84.3% — confirming that domain-specific fine-tuning is essential for medical imaging. Fine-tuning without the attention gate improves to 93.1%. Adding the spatial attention gate improves to 95.2%. The complete model with MixUp augmentation achieves 96.4%, confirming that each component contributes meaningful performance gains.

D. Grad-CAM Radiologist Validation

Three radiologists evaluated 150 Grad-CAM heatmap outputs. The clinical relevance score averaged 4.3/5.0 (SD=0.61), with radiologists confirming that the model attention regions correctly highlighted periosteal reactions, cortical destruction, medullary involvement, and soft-tissue extension in 91.3% of malignant cases. False attention regions were most common in Chondrosarcoma cases (12.4% mismatch), consistent with the class-level classification difficulty. Radiologist feedback identified Grad-CAM explainability as the most clinically valuable system feature, enabling trust-calibrated use of AI predictions.

E. Inference Time and Deployment

The trained model is deployed as a Flask-based web osteosarcoma MRI segmentation, and Liu et al. [9] applied 3D U-Net for bone metastasis detection on MRI. Neither study incorporated multi-class classification across multiple primary bone cancer types or provided visual explainability outputs for radiologist validation. Our work addresses these gaps by providing a unified multi-class classification system across both X-Ray and MRI modalities with integrated Grad-CAM explainability validated by clinical experts.

III. PROPOSED METHODOLOGY

A. Dataset Preparation

The dataset comprises 3,200 annotated bone imaging samples assembled from three sources: The Cancer Imaging Archive (TCIA) public repository, the MURA musculoskeletal X-Ray dataset from Stanford University, and an institutional dataset of 320 de-identified patient cases from affiliated hospitals with confirmed histopathological diagnoses and IRB ethics approval (Ref: SJRIT-IEC-2025-04). The five-class distribution is: Osteosarcoma (850 images), Ewing Sarcoma (620), Chondrosarcoma (580), Benign Bone Tumor (650), and Normal Bone (500). The imaging modality mix includes X-Ray (45%), MRI T1/T2 sequences (40%), and CT scout images (15%). The dataset is partitioned into training (70%), validation (15%), and test (15%) sets using stratified sampling to ensure class balance.

B. Preprocessing and Augmentation

All images undergo a standardized preprocessing pipeline: resizing to 224x224 pixels (EfficientNet-B4 input standard), normalization using ImageNet channel mean and standard deviation, and modality-specific preprocessing — CLAHE (Contrast Limited Adaptive Histogram Equalization) for X-Ray images, Z-score intensity normalization for MRI sequences, and Hounsfield Unit windowing (W:2000, L:300) for CT images. Data augmentation is applied exclusively to the training set using: random horizontal and vertical flipping, rotation (plus or minus 30 degrees), elastic deformation, Gaussian noise addition, brightness and contrast jitter, and MixUp augmentation — expanding the effective training dataset to 15,000 images.

C. Model Architecture

The proposed architecture uses EfficientNet-B4 (pre-trained on ImageNet-21k) as the backbone feature extractor with its final classification head replaced by a custom module comprising: a Spatial Attention Gate that generates channel-wise and spatial attention maps to focus on diagnostically relevant image regions, Global Average Pooling (GAP) to reduce spatial dimensions, a Dropout layer (rate=0.4) for regularization, a Dense layer (256 units, ReLU activation), and a final Softmax output layer with five units corresponding to the five diagnostic classes. The model is trained end-to-end using the Adam optimizer (learning rate=1e-4, weight application. Average inference time per image is 1.8 seconds (including image preprocessing, model forward pass, and Grad-CAM heatmap generation), well within the 5-second target for clinical workflow integration. The web interface allows DICOM and standard image format uploads, displays the predicted class with confidence scores, and renders the Grad-CAM overlay for radiologist review.

V. CONCLUSION AND FUTURE WORK

This paper presented an AI-based Bone Cancer Detector leveraging a fine-tuned EfficientNet-B4 CNN with spatial attention augmentation for five-class classification of bone lesions from X-Ray and MRI medical images. Trained on a curated 3,200-image dataset and evaluated on a held-out test set, the proposed model achieves 96.4% accuracy, 95.8% sensitivity, 97.1% specificity, and AUC-ROC of 0.982 — substantially outperforming VGG16, ResNet-50, InceptionV3, and DenseNet-121 baselines.

Grad-CAM heatmap visualization, validated by three independent radiologists with an average clinical relevance score of 4.3/5.0, provides the interpretability essential for clinical trust and adoption. The ablation study confirms that each architectural component — fine-tuning, spatial attention, and MixUp augmentation — contributes measurable performance improvements. The Flask-based web deployment with 1.8-second inference time demonstrates practical clinical feasibility.

Future work will pursue four directions: (1) extending the system to CT scan volumes using 3D CNN architectures for volumetric tumor segmentation alongside classification, (2) implementing federated learning to enable privacy-preserving collaborative model training across multiple hospitals without sharing raw patient imaging data, (3) integrating clinical metadata including patient age, anatomical location, and serum biomarkers as auxiliary inputs to improve classification accuracy for ambiguous cases, and (4) prospective clinical validation in collaboration with district hospital radiology departments in Dhule to assess real-world diagnostic impact.

REFERENCES

A. Indian Council of Medical Research (ICMR). National Cancer Registry Programme Report 2020. ICMR, New Delhi, 2020.

B. C. E. Von Schacky et al., Multitask deep learning for segmentation and classification of primary bone tumors on radiographs, Radiology, vol. 301, no. 2, pp. 398-406, 2021.

C. G. Litjens et al., A survey on deep learning in medical image analysis, Medical Image Analysis, vol. 42, pp. 60-88, 2017.

D. K. Sampath, S. Rajagopal, and A. Chintanpalli, A comparative analysis of CNN-based deep learning architectures for early diagnosis of bone cancer using CT images, Scientific Reports, vol. 14, p. 2144, 2024.

E. A. Sharma, D. P. Yadav, H. Garg, M. Kumar, B. Sharma, and D. Koundal, Bone cancer detection using feature extraction based machine learning model, Computational and Mathematical Methods in Medicine, vol. 2021, Art. ID 7433186, 2021.

F. D. M. Anisuzzaman et al., A deep learning study on osteosarcoma detection from histological images, Biomedical Signal Processing and Control, vol. 69, p. 102931, 2021.

decay=1e-5), categorical cross-entropy loss, and cosine annealing learning rate scheduling for 60 epochs with early stopping (patience=8).

D. Grad-CAM Explainability

Gradient-weighted Class Activation Mapping (Grad-CAM) is applied to the final convolutional block of the EfficientNet-B4 backbone to generate class-discriminative heatmaps. These heatmaps are overlaid on the original medical images using a jet colormap, highlighting the spatial regions that most strongly influence the model classification decision. Three independent radiologists (with 8, 12, and 15 years of musculoskeletal imaging experience) evaluated 150 randomly sampled Grad-CAM outputs in a blinded review, rating the clinical relevance of highlighted regions on a 5-point Likert scale.

G. C. E. Von Schacky et al., Multitask deep learning for segmentation and classification of primary bone tumors on radiographs, Radiology, vol. 301, no. 2, pp. 398-406, 2021.

H. J. Wu et al., A residual fusion network for osteosarcoma MRI image segmentation in developing countries, Computational Intelligence and Neuroscience, vol. 2022, Art. ID 9467264, 2022.

I. X. Liu et al., Detection and segmentation of pelvic bones metastases in MRI images for patients with prostate cancer based on deep learning, Frontiers in Oncology, vol. 11, p. 773299, 2021.

J. M. Tan and Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in Proc. ICML, 2019, pp. 6105-6114.

K. R. R. Selvaraju et al., Grad-CAM: Visual explanations from deep networks via gradient-based localization, in Proc. ICCV, 2017, pp. 618-626.

ISSN: Applied For (2026) | DOI: 10.XXXXX/IJVAST.2026.XXXXX | Open Access — CC BY 4.0 License