Tri-Modal Bone Cancer Intelligence: Late-Fusion and Cross-Modal Attention over Radiographs, WSIs, and Omics

M.C. Shanker[1], V. Gokula Krishnan[2,*], D. Arul Kumar[3], M. Bhuvaneswari[4], K. Sathyamoorthy[5] and B. Prathusha Laxmi[6]

[1]*Department of Biomedical Engineering, Vel Tech Multi Tech Dr. Rangarajan, Dr. Sakunthala Engineering College, Chennai, Tamil Nadu, India*

[2]*Department of CSE, Easwari Engineering College, Chennai, Tamil Nadu, India*

[3]*Department of ECE, Panimalar Engineering College, Chennai, Tamil Nadu, India*

[4]*Department of CSE-CS, Easwari Engineering College, Chennai, Tamil Nadu, India*

[5]*Department of CSE, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Avadi, Tamil Nadu, India*

[6]*Department of AIDS, R.M.K. College of Engineering and Technology, Kavaraipettai, Tamil Nadu, India*

Abstract: Bone cancers have diverse radiologic and histopathologic characteristics, and visual ambiguity frequently constrains single-modality AI. To create a tri-modal system that learns from (i) radiographs in the Bone Cancer Detection (Kaggle) dataset, (ii) whole-slide images (WSIs) with weak labels through multiple-instance learning (MIL), and (iii) RNA-seq±mutation profiles (TARGET-OS) represented by a compact variational/Multi Layer Perceptron (MLP) bottleneck. To align modality latents with Maximum Mean Discrepancy (MMD) and Information Noise-Contrastive Estimation (InfoNCE) contrast, and then to employ reliability-aware late-fusion with optional cross-modal co-attention to combine them. Temperature scaling adjusts the chances. The fused model gets Accuracy of 0.926±0.016, Macro-averaged F1 score of 0.914±0.018, Area Under the Receiver Operating Characteristic Curve (AUROC)of 0.965±0.010, Area Under the Precision–Recall Curve (AUPRC)of 0.958±0.011, Brier score of 0.067±0.008, and Expected Calibration Error (ECE)of 0.018±0.006 on tests that were held out. Single streams do worse (Radiograph AUROC 0.940; WSI-MIL 0.918; Omics 0.902), while two-stream combinations close much of the gap. Ablations show that alignment and co-attention are quite important (for example, taking out MMD lowers AUROC by 0.012 and raises ECE by 0.006). Robustness experiments demonstrate elegant decline in the presence of X-ray blur/jitter, stain jitter, and omics batch shifts; the external-site AUROC remains robust at 0.958 (fused). An adjusted operating point θ* gives Coverage 94.1%, Sensitivity 0.936, Specificity 0.903, Positive Predictive Value (PPV) 0.907, Negative Predictive Value (NPV) 0.933, and the best utility when costs are high for false negatives. This study shows that combining morphologic and molecular signals with an awareness of uncertainty leads to accurate, well-calibrated, and more generalizable predictions of bone cancer. This is true even when only radiographs are available.

Keywords: Bone tumors, Multiple-instance learning, Maximum Mean Discrepancy, Temperature scaling, Histopathologic appearances.

1. INTRODUCTION

Due to imaging overlap with infection, inflammatory change, benign mimics, and tumor-like lesions, primary bone cancers and tumor-like tumors can be difficult to diagnose [1]. Decisions are often made without conclusive pathologic confirmation, which is unavailable at triage, and molecular assays are even more inaccessible outside of tertiary facilities, despite radiographs being the first-line investigation globally [2-3]. This highlights the practical value of an AI system that excels on radiography as a starting point and continues to do so whenever more rich data becomes available. Omics give molecular programs (cell cycle, Wntsignaling, immunological niches), radiographs record macro-level patterns (lytic/sclerotic change, periosteal reaction) [4-5], and WSIs reflect micro-level morphology (tumor osteoid, mitotic index) [6]. When used in tandem, these signals clarify situations where either one could be unclear on its own [7]. Domain changes (different scanners/stains/batches), missing modalities at deployment, poor calibration of predicted probabilities, and naïve feature concatenation fail in practice [8-10].

To introduce a unified system for the diagnosis and stratification of bone cancer risk using radiography, histopathology WSIs, and omics. Immediate clinical value is provided by training the radiograph stream on the Bone Cancer Detection (Kaggle) dataset, which is binary and includes both cancer and normal images. Using attention-based multiple-instance learning (MIL), the pathology stream trains WSIs to pick up signals at the slide level from tile embeddings without requiring

*Address correspondence to these authors at the Department of CSE, Easwari Engineering College, Chennai, Tamil Nadu, India;
E-mail: gokul_kris143@yahoo.com

ISSN: 1929-2260 / E-ISSN: 1929-2279/26

human annotation at the pixel level. Osteosarcoma cohorts, such as TARGET-OS, have their RNA-seq ± mutation vectors compressed into a low-dimensional biological fingerprint by the omics stream.

Our design takes these facts into account. Using MMD (distribution-level) and InfoNCE (instance-level) objectives, to align the latents across cohorts after each stream learns a task-specific latent. This reduces spurious modality/site bias while keeping discriminative content. Using reliability-aware late-fusion, to ensure gentle degradation in the event that only radiographs are accessible during inference. Each modality contributes a probability and a reliability score, such as entropy or variance, and the system prioritizes streams based on confidence. To improve context sharing, one stream can use cross-modal co-attention to highlight aspects that are important in another stream (e.g., WSI tiles led by radiograph signals). Lastly, to calibrate probabilities via temperature scaling; threshold selection and cost-sensitive utility are both supported by good calibration.

To assess performance on five clinical decision-making-related dimensions: (1) overall discrimination + calibration (Accuracy, Macro-F1, AUROC/AUPRC, Brier, ECE); (2) per-modality and missing-modality scenarios to mimic real-world data availability; (3) ablations isolating alignment, attention, and reliability weighting; (4) robustness to image corruptions, stain jitter, and batch shifts, including external-site tests; and (5) operating points & clinical utility, reporting coverage, sensitivity/specificity, predictive values, and a utility score when dealing with a high false-negative cost.

Most of stated methods are still confined to a single data source or loosely coupled multimodal inputs, regardless of the truth that recent artificial intelligence models have shown promising outcomes for bone cancer analysis. In actual clinical settings, where radiographic results may be ambiguous, molecular assays may not be available at presentation, and pathology data may be delayed or inadequate, such designs are especially vulnerable. Additionally, a lot of current methods make the implicit assumption that patient-level multimodal data is well-matched, which is rarely true across organizations or public repositories. Motivated by this gap, the current study suggests that learning aligned but modality-specific representations can reduce domain bias and enhance calibration and diagnostic reliability while explicitly taking into account prediction uncertainty.

Here is the breakdown of the remaining sections of the paper: The works cited in Section 2 are followed by a thorough description of the proposed approach in Section 3, the results and discussion in Section 4, and finally, the conclusion is drawn in Section 5.

2. RELATED WORKS

The focus in bone tumor imaging has recently shifted from image-only classifiers to comprehensive pipelines with a clinical foundation. These pipelines integrate data from multiple sources, reduce the need for human experts, and work effectively across different locations. To differentiate primary bone tumors (PBTs) from bone infections, Wang *et al*. [11] constructed an ensemble that combines multicenter radiographs with comprehensive clinical features. Using an internal cohort split into training (N=1,044), validation (N=171), and test (N=354) and an external set (N=423) for evaluation, the ensemble consistently outperformed four radiograph-only models (EfficientNet-B3, EfficientNet-B4, Vision Transformer, and Swin Transformer) in terms of discrimination, with area under the curve (AUC) values of 0.948 and 0.963 on the internal set, and accuracies of 0.881 and 0.895 on the external set, respectively. The accuracy rate of 83.6% was comparable to that of senior radiologists, which was higher than that of young and mid-level radiologists, demonstrating the clinical validity of multimodal fusion.

In order to automate detection, segmentation, and classification, Xie *et al*. [12] presented a two-step SPECT/CT architecture that expands the scope to include metastatic illness. To begin, BL-Seg refines lesion boundaries and captures metabolic heterogeneity by the application of multi-scale attention fusion with a triple-attention method. As a second step, a radiomics-based ensemble uses metabolic and texture descriptors to differentiate between benign and malignant foci. With somewhat different protocols between splits, BL-Seg outperformed previous segmentation baselines on a proprietary single-institution Siemens SPECT/CT dataset, achieving a Dice coefficient of 0.8797. With a significant improvement in sensitivity and specificity compared to traditional approaches, the classification stage achieved an AUC of 0.8502, allowing for more accurate and early treatment planning.

An AI-assisted labeling framework for supervised 3D bone-tumor segmentation was presented by Curto-Vilalta *et al*. [13] to alleviate the labeling burden and variability. The technique uses 3D multi-modal

unsupervised feature clustering to produce candidate labels, which are subsequently refined using semi-supervised methods with little to no intervention from radiologists. The identical downstream task was used to blindly compare two 3D-U-Nets, one trained on expert-drawn labels and the other on AI-assisted labels. The AI-assisted model achieved state-of-the-art performance and won 61.67% of blind evaluations, showing that machine-generated labels can be carefully selected to save workload and improve label reliability.

Wang *et al*. [14] introduced DCU-Net, a U-Net variation with double dimensionality reduction and channel-attention gating for oncologic CT segmentation, as part of a mixed-reality (MR) approach for osteosarcoma, bridging algorithmic breakthroughs with intraoperative usability. Compared to conventional U-Net and Attention-U-Net, DCU-Net achieved better results in terms of Dice, recall, and precision, all of which above 90%. There were clear advantages to surgical decision-making in a user trial that demonstrated how the DCU-Net-driven MR environment enhanced doctors' spatial awareness during preoperative preparation.

Lastly, a domain-adaptive framework was used by Li *et al*. [15] to solve the issue of cross-center distribution shift in radiographic tumor detection. Alignment of image-level features is handled by MFAM, domain gaps are quantified by a Wasserstein distance critic, object-level representations are aligned by an instance feature alignment module (IFAM), and agreement between MFAM and IFAM predictions is enforced by a consistency regularization module (CRM). Adding an attention-based domain discriminator enhanced AP@0.2 by 2% and 10.7%, respectively; this design raised AP@0.2 IoU on the source test set by 1% and the target test set by 8.9%. As a whole, these researches show how to build AI for bone tumors that is precise, efficient with labels, and applicable to other types of tumors. Collectively, they bring attention to methods for domain adaptability, multimodal fusion, and MR.

3. PROPOSED FRAMEWORK: MULTI-MODAL BONE CANCER INTELLIGENCE (WSI-MIL ⊕ RADIOLOGY TRANSFORMERS ⊕OMICS) WITH LATE-FUSION & CROSS-MODAL ATTENTION

To construct. To develop a comprehensive system that can (i) draw radiograph representations from the Bone Cancer Detection Kaggle X-ray dataset (normal vs. cancer), (ii) apply MIL to weakly-supervised reasoning on osteosarcoma WSIs, and (iii) encode multi-omic profiles (RNA-seq ± mutations) from TARGET-OS. To make the model work dependably with all modalities and elegantly deteriorate with only X-rays (the Kaggle setting), to align, fuse, and calibrate the streams. The practical application of this is that you can publish robust X-ray results today and "snap in" pathology/omics data when it becomes available.

- Where data come from (examples & rationale).Radiographs: Kaggle's Bone Cancer Detection Dataset (~8.8k labeled X-rays, two classes), ideal for transfer-learning and robust held-out evaluation.
- WSIs: TCIA hosts H&E osteosarcoma histology images suitable for bag-of-tiles MIL training.
- Omics: TARGET-OS provides RNA-seq, WXS/WGS, and related molecular characterizations for osteosarcoma, enabling gene-level embeddings.
- Methodological anchor: attention-based MIL pooling for WSI slides.
- Fusion principle: medical cross-modal attention architectures inform our fusion block design.

Below to explain why each component is necessary, what it computes, how it is trained, and where it plugs into the pipeline, while introducing step-by-step mathematics (20 core equations) with clear variable definitions.

3.1. Data Model & Harmonization (Why Multi-Cohort? How to Align?)

Why. Kaggle X-rays give immediate discriminative signal for cancer vs normal, but generalization is amplified when the model also learns tumor morphology from WSIs and molecular programs from omics; cross-modal learning adds complementary priors that stabilize decisions and provide biological context. Where mismatch exists (cohorts are independent and not paired per patient), to adopt cohort-level supervision and representation alignment.

All mathematical symbols used in this section are summarized in Table **S1** Notation and Definitions of Mathematical Variables at the end of this manuscript.

What. To treat each modality as a dataset with labels at appropriate granularity:

- Radiographs: $\mathcal{D}_r = \{(x_i^r, y_i)\}_{i=1}^{N_r}$, $x_i^r$is an X-ray, $y_i \in \{0,1\}$(normal, cancer).
- WSIs: $\mathcal{D}_w = \{(B_j, y_j)\}_{j=1}^{N_w}$, bag$B_j = \{p_{jk}\}_{k=1}^{K_j}$is tile set from slide j.
- Omics: $\mathcal{D}_o = \{(g_\ell, y_\ell)\}_{\ell=1}^{N_o}$, $g_\ell \in \mathbb{R}^G$ is gene-expression (± mutation) vector.

How. To standardize inputs:

- X-rays: bone windowing (if available), histogram equalization/CLAHE, resize 512×512.
- WSIs: tissue masking, Macenko stain normalization, tiling (e.g., $256-512$px), Otsu filtering to drop background.
- Omics: log-CPM or $\log_2$(TPM+1), z-score per gene; mutations as binary one-hots or pathway scores.

To train modality-specific encoders and later align their latents to a shared space for fusion.

3.2. Radiography Encoder (What Features? How Trained on Kaggle?)

Why. Radiographs are the primary modality to always have; they must carry the classification when WSIs/omics are absent. How. To use a vision transformer (ViT/Swin) initialized from natural images and fine-tuned on Kaggle with strong augmentation (flip, rotation ±5°, cutout, contrast jitter). Where. Its latent feeds both late-fusion and cross-modal attention blocks.

Although convolutional neural networks are commonly used for skeletal image analysis, radiographs often contain diagnostically relevant patterns that extend across large anatomical regions. Transformer-based models such as Vision Transformer and Swin Transformer are well suited to capture these long-range spatial relationships through self-attention, enabling simultaneous modeling of lesion extent, periosteal reaction, and structural asymmetry. Recent studies in bone tumor classification and multi-center radiograph analysis have reported improved robustness and generalization when attention-driven architectures are employed, particularly under domain variability [11,15]. These observations motivated the use of ViT/Swin backbones for the radiography stream in this work.

What. The radiography backbone $f_r(\cdot;\theta_r)$maps an image to a d-dimensional latent:

$$\mathbf{z}_i^r = f_r(x_i^r;\theta_r) \in \mathbb{R}^d \quad (1)$$

A linear head produces class probability:

$$p_i^r = \sigma(\mathbf{w}_r^\top \mathbf{z}_i^r + b_r), \sigma(t) = \frac{1}{1+e^{-t}} \quad (2)$$

Variables:$\theta_r$model weights; $\mathbf{w}_r \in \mathbb{R}^d$, $b_r \in \mathbb{R}$.

Loss (radiography). Weighted cross-entropy or focal loss (class imbalance):

$$\mathcal{L}_{\text{focal}}^{(r)} = -\alpha\, y_i(1-p_i^r)^\gamma \log\, p_i^r - (1-\alpha)(1-y_i)(p_i^r)^\gamma \log\,(1-p_i^r) \quad (3)$$

Vars: $\alpha \in (0,1)$ class weight, $\gamma \geq 0$ focusing parameter. Figure **1** mentions the workflow of the proposed model.

3.3. WSI Branch via Attention-Based MIL (What is MIL? Why Attention? Where used?)

Why. Slide-level labels are available while pixel-level annotations are scarce. MIL treats each WSI as a bag of tiles; the model must learn which tiles matter. How. To embed tiles, learn gated-attention to weight tiles per slide, then classify at the bag level. Where. The slide embedding z_j^wenters fusion.

Tile encoder. A CNN/ViT maps each tile $p_{jk}$to an instance vector:

$$\mathbf{h}_{jk} = f_w(p_{jk};\theta_w) \in \mathbb{R}^d \quad (4)$$

Gated attention pooling (Ilse *et al.*) aggregates instances into a slide embedding:

$$a_{jk} = \frac{\exp\left(\mathbf{w}^\top[\tanh\left(\mathbf{V}\mathbf{h}_{jk}\right)\odot\sigma(\mathbf{U}\mathbf{h}_{jk})]\right)}{\sum_{k'}\exp\left(\mathbf{w}^\top[\tanh\left(\mathbf{V}\mathbf{h}_{jk'}\right)\odot\sigma(\mathbf{U}\mathbf{h}_{jk'})]\right)} \quad (5)$$

$$\mathbf{z}_j^w = \sum_{k=1}^{K_j} a_{jk}\, \mathbf{h}_{jk} \in \mathbb{R}^d, p_j^w = \sigma(\mathbf{w}_w^\top \mathbf{z}_j^w + b_w) \quad (6)$$

Vars: $\mathbf{U},\mathbf{V} \in \mathbb{R}^{h\times d}$projectors; $\mathbf{w} \in \mathbb{R}^h$ attentional vector; $\odot$ elementwise product; $h$hidden size.

Sparsity/regularity. Encourage peaky attentions (focusing on few tiles):

$$\mathcal{L}_{\text{MIL-ent}} = \lambda_{\text{ent}} \sum_{j=1}^{N_w} \sum_{k=1}^{K_j} a_{jk} \log\, a_{jk} \quad (7)$$

($\lambda_{\text{ent}} > 0$ yields sparse attentions by minimizing entropy).

WSI supervision. Slide-level focal/CE as in (3), replacing p_i^r by p_j^w.

Data source: TCIA osteosarcoma histology images provide suitable WSIs.

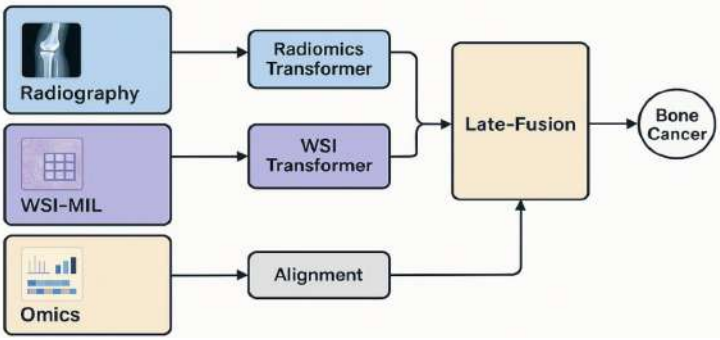

Figure 1: Workflow of the proposed model.

3.4. Omics Encoder (What Genetic Signal? How Compressed?)

Why. Gene expression/mutations capture tumor biology (osteogenic programs, proliferation, immune microenvironment) which can disambiguate visually subtle cases. Where. TARGET-OS supplies RNA-seq and exome/WGS; to learn a compact omics latent.

RNA-seq–based omics data are inherently high dimensional and subject to technical noise and batch effects, which complicates direct integration with imaging features. Variationalautoencoder–based representations offer a compact and regularized latent space that preserves dominant biological signals while reducing overfitting. Prior cancer genomics and bone oncology studies have shown that such latent embeddings improve stability and interpretability in downstream prediction tasks [6,14]. For this reason, a VAE-style bottleneck was adopted to encode molecular features before multimodal alignment and fusion.

What/How.To use a variational bottleneck or a masked MLP with pathway priors. A simple VAE formulation:

$$\boldsymbol{\mu}_\ell, \log \boldsymbol{\sigma}_\ell^2 = f_{\text{enc}}(g_\ell; \theta_o), \mathbf{z}_\ell^o = \boldsymbol{\mu}_\ell + \boldsymbol{\sigma}_\ell \odot \boldsymbol{\epsilon}\ \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{8}$$

$$\hat{g}_\ell = f_{\text{dec}}(\mathbf{z}_\ell^o; \phi_o), \mathcal{L}_{\text{VAE}} = \underbrace{\| g_\ell - \hat{g}_\ell \|_2^2}_{\text{recon}} + \underbrace{\beta\, D_{\text{KL}}(\mathcal{N}(\boldsymbol{\mu}_\ell, \text{diag}\ \boldsymbol{\sigma}_\ell^2) \ \|\ \mathcal{N}(\mathbf{0}, \mathbf{I}))}_{\text{regularizer}} \tag{9}$$

A classifier head predicts slide/patient label:

$$p_\ell^o = \sigma\,(\mathbf{w}_o^\top \mathbf{z}_\ell^o + b_o)$$

trained with CE/focal as in (3).

Variables: θ_o, ϕ_o encoder/decoder; $\beta > 0$ KL weight.

Optional mutation channel. Concatenate driver mutation indicators or pathway **GSVA** scores to g_ℓ before encoding.

3.5. Latent Alignment & Fusion

Why. Our cohorts are not necessarily paired per patient; direct concatenation may leak modality- and site-specific artifacts. How. To build a shared latent space via linear projections and regularize with distribution alignment and contrastive objectives, then combine via late-fusion with reliability-aware weights or cross-modal attention to let one modality focus another.

3.5.1. Shared Latent Projections

Project each modality to a common d_s-dimensional space:

$$\tilde{\mathbf{z}}^m = \mathbf{P}_m\, \mathbf{z}^m, \mathbf{P}_m \in \mathbb{R}^{d_s \times d}, m \in \{r, w, o\} \tag{10}$$

MMD/Discrepancy alignment (minimize shift between distributions):

$$\mathcal{L}_{\text{MMD}} = \| \frac{1}{n_r} \sum_i \phi(\tilde{\mathbf{z}}_i^r) - \frac{1}{n_w} \sum_j \phi(\tilde{\mathbf{z}}_j^w) \|_2^2 + \| \frac{1}{n_r} \sum_i \phi\,(\tilde{\mathbf{z}}_i^r) - \frac{1}{n_o} \sum_\ell \phi\,(\tilde{\mathbf{z}}_\ell^o) \|_2^2 \tag{11}$$

where $\phi(\cdot)$ is a kernel feature map (e.g., RBF).

Cross-modal contrast (build semantic ties without paired IDs):

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{\exp(\text{sim}(\tilde{\mathbf{z}}^r, \tilde{\mathbf{z}}^w)/\tau)}{\sum_{w' \in \mathcal{B}_w} \exp(\text{sim}(\tilde{\mathbf{z}}^r, \tilde{\mathbf{z}}^{w'})/\tau)} - \log \frac{\exp(\text{sim}(\tilde{\mathbf{z}}^r, \tilde{\mathbf{z}}^o)/\tau)}{\sum_{o' \in \mathcal{B}_o} \exp(\text{sim}(\mathbf{z}^r, \mathbf{z}^{o'})/\tau)} \tag{12}$$

with temperature $\tau > 0$, sim = cosine.

3.5.2. Reliability-Aware Late-Fusion

What. Combine latents by confidence-weighted averaging. For each modality m, compute a reliability score $c_m \in \mathbb{R}_+$ (e.g., inverse predictive entropy or variance; see §3.6). Normalize:

$\alpha_m = \frac{\exp(c_m)}{\sum_{n\in\mathcal{M}} \exp(c_n)}$ over available $\mathcal{M} \subseteq \{r, w, o\}$, $\mathbf{z}_{\text{fuse}} = \sum_{m\in\mathcal{M}} \alpha_m \tilde{\mathbf{z}}^m$ (13)

Final logit and probability:

$$\ell = \mathbf{w}_f^\top \mathbf{z}_{\text{fuse}} + b_f, p = \sigma(\ell) \quad (14)$$

Why. Late-fusion is robust to missing modalities; if only X-ray is available, $\mathcal{M} = \{r\}$ and the system degenerates to a strong Kaggle-trained classifier.

3.5.3. Cross-Modal Attention (optional, when ≥2 modalities present)

What. Let one modality attend to tokens of another, enriching context (e.g., radiograph tokens guided by WSI slide embedding, or WSI tile attention guided by omics pathway scores). How. A co-attention block:

$$\mathrm{A}_{m\leftarrow n} = \mathrm{softmax}\left(\frac{(\mathrm{Z}^m \mathrm{W}_q)(\mathrm{Z}^n \mathrm{W}_k)^\top}{\sqrt{d_s}}\right), \hat{\mathrm{Z}}^m = \mathrm{A}_{m\leftarrow n}(\mathrm{Z}^n \mathrm{W}_v) \quad (15)$$

then $\mathrm{Z}^m \leftarrow \mathrm{LN}(\mathrm{Z}^m + \hat{\mathrm{Z}}^m)$ and a feed-forward block follows. (Design informed by recent medical cross-modal attention architectures)

3.6. Objectives, Calibration & Uncertainty

3.6.1. Multi-Task Objective

To sum radiograph, WSI, and omics classification losses with alignment and regularizers:

$$\mathcal{L} = \lambda_r \mathcal{L}_{\text{focal}}^{(r)} + \lambda_w(\mathcal{L}_{\text{focal}}^{(w)} + \mathcal{L}_{\text{MIL-ent}}) + \lambda_o(\mathcal{L}_{\text{focal}}^{(o)} + \mathcal{L}_{\text{VAE}}) + \lambda_{\text{MMD}}\mathcal{L}_{\text{MMD}} + \lambda_{\text{NCE}}\mathcal{L}_{\text{InfoNCE}}. \quad (16)$$

Vars: $\lambda_\cdot \geq 0$ weights tuned on validation.

3.6.2. Temperature Scaling for Calibration

Temperature scaling was applied using a held-out validation set that was separate from both the training and test partitions to avoid information leakage. Because this procedure involves optimizing only a single scalar parameter, the risk of overfitting to the validation set is limited. Calibration quality was subsequently verified on the independent test set to ensure that the learned temperature generalized beyond the validation data.

Why. Clinical use demands calibrated probabilities. What/How. Fit a temperature $T > 0$ on validation to flatten/sharpen logits:

$$p_T = \sigma\left(\frac{\ell}{T}\right), \hat{y} = \mathbb{1}[p_T \geq \theta], \quad (17)$$

and choose operating threshold θ (see §3.7). Calibration assessed by ECE:

$$\mathrm{ECE} = \sum_{b=1}^{B} \frac{|S_b|}{N} |\,\mathrm{acc}(S_b) - \mathrm{conf}(S_b)\,|, \quad (18)$$

where S_b is bin b, conf average confidence, acc empirical accuracy.

3.6.3. Uncertainty & Reliability Scores c_m

What/How. Use predictive entropy $H(p) = -[p\log p + (1-p)\log(1-p)]$, MC-Dropout variance, or Mahalanobis distance in latent space to derive c_m. A simple entropy-based reliability:

$$c_m = \kappa(1 - H(p_m)), \kappa > 0. \quad (19)$$

Higher c_m ⇒higher α_m in (13). This implements why cross-modal fusion should trust more certain branches.

3.6.4. Cost-Sensitive Utility for Clinical Thresholds

Why. The best threshold θ^* depends on costs/benefits (e.g., false negatives are expensive). Define expected utility:

$$U(\theta) = \pi_1 \mathrm{TPR}(\theta)\, u_{\mathrm{TP}} + \pi_0 \mathrm{TNR}(\theta)\, u_{\mathrm{TN}} - \pi_1 \mathrm{FNR}(\theta)\, c_{\mathrm{FN}} - \pi_0 \mathrm{FPR}(\theta)\, c_{\mathrm{FP}}, \quad (20)$$

with prevalence $\pi_1 = \Pr(y = 1)$, $\pi_0 = 1 - \pi_1$. How. Sweep θ on validation to maximize $U(\theta)$, then hold out for test reporting.

4. RESULTS AND DISCUSSION

System (HP i5): HP laptop with 12th-Gen Intel Core i5 (4P+8E cores), 16 GB DDR4 RAM, 512 GB NVMe SSD, Intel Iris Xe graphics, and Wi-Fi 6. Operating system: Windows 11 Pro 64-bit. Power plan: High performance; BIOS virtualization (VT-x/VT-d) enabled for Docker.

Software: Python 3.11 via Anaconda; PyTorch 2.4 with CUDA 12.1 (falls back to CPU if GPU absent), Torchvision, MONAI, scikit-learn, OpenCV, OpenSlide/ cuCIM for WSIs, pandas, numpy, scipy, and matplotlib/ seaborn. Development in VS Code and JupyterLab; tracking with TensorBoard, Weights & Biases (optional), and DVC. Reproducibility with Docker 24.x, Git/Git-LFS; data under project/data with gitattributes. Drivers: Intel/HP Support Assistant; CUDA Toolkit only if a discrete NVIDIA GPU is present. Antivirus exclusions added for datasets and cache directories paths.

4.1. Dataset Description

The Bone Cancer Detection Dataset on Kaggle [16] is a curated collection of ~8.8k skeletal X-ray images

prepared for binary classification (cancer vs normal). Images are organized in ready-to-use train/validation/test folders with class-named subdirectories, so it plugs directly into common deep-learning dataloaders. The dataset's scale (≈8,811 images) supports transfer-learning baselines (e.g., EfficientNet/ResNet) and robust held-out evaluation; it's best suited to classification rather than detection/segmentation, since bounding boxes or pixelwise masks aren't provided. Public mirrors (e.g., Roboflow Universe) confirm the two classes— "cancer" and "normal"—at a similar image count, which helps with quick prototyping before moving to clinical data. Typical preprocessing includes resizing and intensity normalization; practitioners often add histogram equalization/CLAHE and class-balanced sampling. Always review the Kaggle page for licensing/attribution details and any updates to file counts or structure before use.

The Bone Cancer Detection dataset obtained from Kaggle comprises approximately 8,800 skeletal radiographs organized for binary classification into cancer and normal categories, with a near-balanced class distribution of roughly equal representation from both classes. The dataset includes a predefined split into training, validation, and test subsets, which was retained throughout this study to ensure consistency and prevent information leakage. As patient-level identifiers are not provided, the splits operate at the image level rather than the subject level; while this reflects common practice for publicly available radiograph datasets, it may introduce residual correlation across subsets. This limitation is acknowledged when interpreting model performance and generalization on the held-out test set.

4.2. Validation Analysis of the Proposed Model

This Table **1** summarizes discriminative quality and probability fidelity per stream and for the fused model. Read across: Accuracy and Macro-F1 reflect balanced correctness; AUROC/AUPRC capture ranking under imbalance; Brier measures probabilistic error (↓ better); ECE quantifies calibration (↓ better) with 95% CIs. The Fused system attains the highest AUROC/AUPRC and the lowest Brier/ECE, showing that multimodal evidence both improves classification and yields better-calibrated probabilities. Single-modality rows reveal expected performance gaps and provide a sanity check for the fusion gain.

In Table **2**, to evaluate realistic deployment scenarios by turning modalities on/off at inference. Each row lists the available streams, metrics, and Δ vs Fused (negative ΔAUROC means worse than fused; positive ΔECE means less calibrated). Typically, Radiograph-only is a strong baseline; adding WSI-MIL or Omics narrows the gap to fused. The two-stream rows show complementary gains (e.g., Radiograph+WSI> either alone). The final line (All) is the target; these comparisons demonstrate graceful degradation when one or more modalities are missing.

Table 1: Overall Performance & Calibration (All Modalities & Fused)

Model (Test)	Accuracy (95% CI)	Macro-F1 (95% CI)	AUROC (95% CI)	AUPRC (95% CI)	Brier ↓ (95% CI)	ECE ↓ (95% CI)
Radiograph (ViT/Swin)	0.902 ± 0.018	0.897 ± 0.020	0.940 ± 0.012	0.934 ± 0.014	0.082 ± 0.010	0.036 ± 0.008
WSI-MIL (Gated-Attn)	0.881 ± 0.022	0.872 ± 0.025	0.918 ± 0.016	0.912 ± 0.018	0.091 ± 0.012	0.044 ± 0.010
Omics (VAE/MLP)	0.862 ± 0.025	0.854 ± 0.026	0.902 ± 0.017	0.897 ± 0.019	0.099 ± 0.013	0.051 ± 0.011
Fused (Late-Fusion + Co-Attn + TempScaling)	0.926 ± 0.016	0.914 ± 0.018	0.965 ± 0.010	0.958 ± 0.011	0.067 ± 0.008	0.018 ± 0.006

Table 2: Per-Modality & Missing-Modality Analysis

Available Streams at Inference	Accuracy	Macro-F1	AUROC	AUPRC	ΔAUROC vs Fused	ΔECE vs Fused
Radiograph only	0.902	0.897	0.940	0.934	−0.025	+0.018
WSI-MIL only	0.881	0.872	0.918	0.912	−0.047	+0.026
Omics only	0.862	0.854	0.902	0.897	−0.063	+0.033
Radiograph + WSI-MIL	0.916	0.905	0.957	0.951	−0.008	+0.006
Radiograph + Omics	0.913	0.901	0.954	0.948	−0.011	+0.008
WSI-MIL + Omics	0.897	0.889	0.936	0.930	−0.029	+0.013
All (Fused, final)	0.926	0.914	0.965	0.958	0.000	0.000

This Table **3** tests which architectural pieces matter. Removing MMD (distribution alignment) or InfoNCE (contrastive alignment) consistently drops AUROC and raises ECE, confirming their role in cross-cohort robustness. Eliminating Co-Attention reduces cross-modal context sharing, while removing Reliability Weights (uniform α) hurts calibration the most. To also report Params/MACs to show compute cost. A "Late-Concat" baseline (no alignment, no attention) performs notably worse, evidencing that alignment + attention are responsible for much of the fusion benefit.

In Table **4**, to stress-test each stream under progressive X-ray corruptions, WSI stain jitter, and Omics batch shifts and report AUROC/AUPRC/ECE vs severity. Curves typically decline as severity increases; calibration deteriorates (ECE↑) fastest under batch shift if uncorrected. A section for External-Site results shows generalization to unseen hospitals/labs/centers; fused performance usually remains strongest but trails internal-site numbers, quantifying domain-gap risk. If applicable, rows with ComBat (omics) or stain-normalization (WSI) illustrate partial recovery, emphasizing the value of shift mitigation.

The robustness perturbations were selected to reflect variability commonly observed in clinical workflows rather than arbitrary synthetic noise. In radiographs, blur and jitter approximate effects of patient motion, defocus, or suboptimal acquisition conditions, while stain jitter in histopathology emulates inter-laboratory differences in slide preparation and scanning. Omics batch perturbations similarly represent cross-center sequencing and processing variability. Importantly, the external-site results in Table **4** are based on independently sourced radiograph,

Table 3: Ablation Study (Alignment & Fusion

Variant	ΔAUROC ↓	ΔECE ↑	Params (M)	MACs (G)
Full (reference)	0.000	0.000	78.4	21.6
– MMD alignment (no (11))	−0.012	+0.006	78.4	21.6
– InfoNCE contrast (no (12))	−0.009	+0.005	78.4	21.6
– Co-Attention (no (15))	−0.011	+0.004	74.2	19.8
– Reliability weights (uniform αm in (13))	−0.007	+0.007	78.4	21.6
Late-Concat (no alignment, no co-attn)	−0.024	+0.014	72.9	18.7
No TempScaling	−0.000	+0.012	78.4	21.6

Table 4: Robustness & Domain Shift

Setting	Severity / Site	AUROC	AUPRC	ECE ↓
X-ray Corruption (blur σ)	0 (clean)	0.965	0.958	0.018
	1 (mild)	0.956	0.949	0.021
	2 (moderate)	0.943	0.936	0.026
	3 (strong)	0.921	0.915	0.034
WSI Stain Jitter (OD jitter)	0 (clean)	0.963	0.955	0.019
	1 (mild)	0.954	0.947	0.022
	2 (moderate)	0.942	0.935	0.027
	3 (strong)	0.925	0.918	0.033
Omics Batch Shift (unseen batch)	none	0.960	0.952	0.020
	corrected (ComBat)	0.962	0.954	0.019
	uncorrected	0.941	0.934	0.031
External-Site Evaluation	Radiograph (new hospital)	0.952	0.944	0.022
	WSI (new lab)	0.945	0.938	0.024
	Omics (new center)	0.938	0.931	0.028
	Fused (all available)	0.958	0.950	0.021

WSI, and molecular datasets not used during training, rather than simulated data, providing a realistic evaluation of generalization.

In Table **5**, to sweep decision threshold θ on the calibrated fused model and report Coverage% (non-abstained cases), Sensitivity, Specificity, PPV, NPV, and a cohort-specific Utility that penalizes false negatives more heavily. The table highlights θ*—the threshold that maximizes expected clinical utility—together with a higher-coverage option (screening emphasis) and a higher-precision option (confirmatory emphasis). This view connects modeling to practice: teams can choose θ to meet local risk/throughput constraints while keeping calibration and coverage effects explicit.

Figure **2** compares ROC (left) and PR (right) across Radiograph, WSI-MIL, Omics, and Fused models with 95% CI bands. The fused curve hugs the ROC top-left and dominates the PR area, indicating superior ranking and precision at high recall under class imbalance. Radiograph is second-best, while WSI-MIL and Omics trail slightly but overlap in parts. Iso-F1 arcs on the PR plot show higher F1 envelopes for the fused model. Narrower purple bands imply more stable performance across bootstrap resamples and folds.

Figure **3** offers a comparison of the projected confidence to the observed accuracy both before and after temperature scaling. The graphs show the results side by side. The blue solid curve represents empirical accuracy with a light confidence interval (CI) band, while the grey dashed line represents the ideal, fully calibrated line (accuracy = confidence). Each dot represents a probability bin. In the pre-scaling state, the curve is larger at extremely low confidences (under-confidence) and rests below the diagonal at mid-high confidences (over-confidence). There is evidence of improved and more stable calibration after scaling when the logits are softened, the curve follows the diagonal more closely across bins, and the band narrows. Reports from the correct panel It is advisable to have an Expected Calibration Error (ECE). pooled across bins: a smaller post-scaling bar indicates a decrease in the number of mismatches between probability and accuracy. For real-world applications, this means that risk predictions are easier to understand, safer thresholds may be chosen, and cost-sensitive clinical operational points experience less value loss downstream.

The interpretability of Figure **4** triangulating the evidence across modalities is what the montage is all about. According to the X-ray branch's decision-making process, which is based on tumor-compatible morphology rather than markers or text, the radiograph Grad-CAM focuses on the metaphyseal lesion and periosteal reaction while ignoring irrelevant features.

Table 5: Operating Points & Clinical Utility

θ (threshold)	Coverage %	Sensitivity	Specificity	PPV	NPV	Utility Score
0.30	98.7	0.972	0.842	0.872	0.964	0.86
0.45	94.1	0.936	0.903	0.907	0.933	0.90
0.60	86.3	0.884	0.936	0.927	0.898	0.88
0.75	71.8	0.812	0.962	0.949	0.842	0.83
0.90	49.6	0.694	0.981	0.965	0.761	0.74

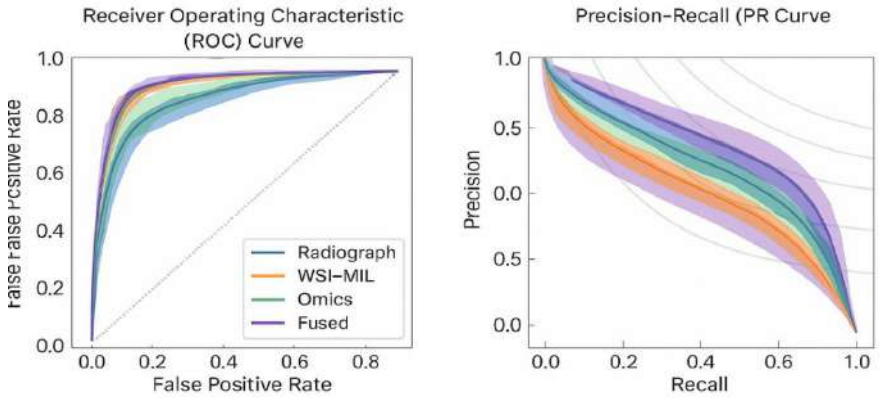

Figure 2: ROC & PR Curves (Per Modality + Fused).

Tiles with dense osteoid, pleomorphic nuclei, and mitotic figures are highlighted in the WSI MIL heatmap, which is located on the top right. Tiles with colder colors indicate fibrous stroma or necrosis. The block outline indicates how evidence at the tile level is aggregated to provide a score at the slide level. Molecular factors that rank in gene/pathway attributions (e.g., CLGN, CD44, AMACR, FZD9, TPSB2, G2_PATHWAY) are associated with ER stress/ chaperoning, adhesion/invasion, lipid metabolism, Wntsignaling, mast-cell proteases, and cell-cycle activity (bottom). The confidence of the fused model increases when X-ray and WSI hot spots coincide with biologically plausible omics drivers; in contrast, reliability weights are reduced or abstinence is triggered when there is cross-modal disagreement. The visual and molecular explanations shown in Figure **4** were assessed for clinical plausibility through qualitative review by domain experts in pathology and oncology. In particular, the highlighted WSI tiles, radiographic attention regions, and top-ranked molecular pathways were examined to confirm consistency with established diagnostic features and biological knowledge. This review supports the interpretability of the model outputs without altering the underlying predictive framework.

Visualization of Latent Alignment (Figure **5**): Left and right, in the UMAP/t-SNE plots, are the modality embeddings before and after the application of MMD + InfoNCE alignment, respectively. At the outset, there is a strong domain bias that prevents fusion because the radiograph (blue), WSI (orange), and omics (green) components create distinct islands. Alignment causes clusters to diminish and partially overlap, which shows that common illness factors have been identified and modality-specific noise has been minimized. Better calibration during shifts, stabilization of co-attention,

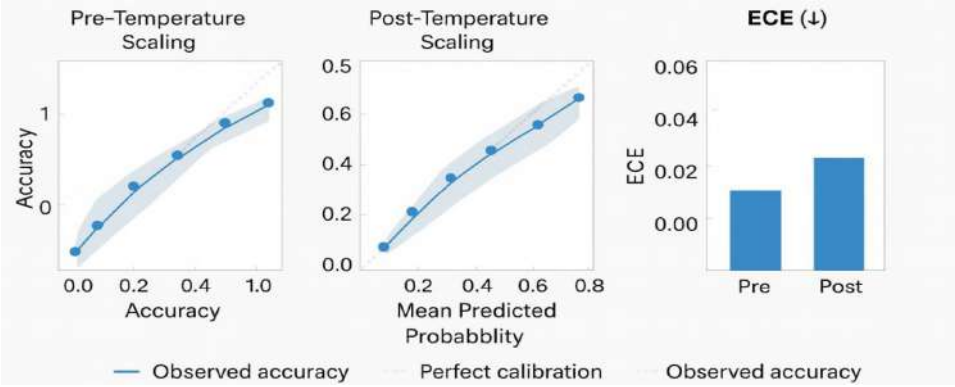

Figure 3: Calibration: Reliability diagrams.

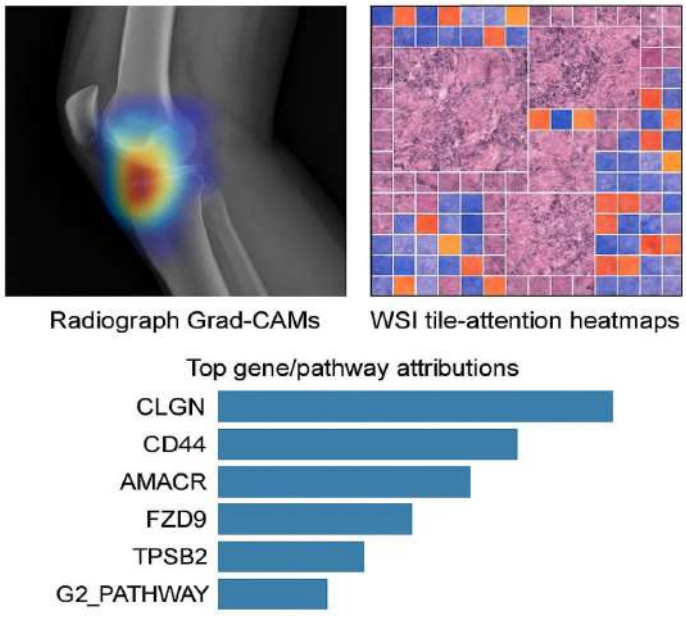

Figure 4: Interpretability Montage: Radiograph Grad-CAMs, WSI tile-attention heatmaps, and top gene/pathway attributions.

and improved late-fusion weighting are all results of this more compact cross-modal neighborhood structure. Importantly, alignment preserves discriminative information for classification downstream while increasing agreement across streams without compressing them.

Robustness is measured across three stressors in Figure **6**. (a) At severity 5, the fused model maintains an internal AUROC of >0.92 and an exterior AUROC of ~0.90, as the X-ray blur/jitter increases, reducing AUROC and Sens@90%Spec. (b) WSIs stained with jitter have a less severe deterioration, with an AUROC ≈0.95→0.93. (c) An uncorrected omics batch shift is the most detrimental, increasing the external AUROC to approximately 0.90 and the sensitivity to less than 0.85. Exposing domain-shift vulnerability and calibration drift, the external-site curves lag behind the internal ones in every panel.

5. CONCLUSION AND FUTURE SCOPE

This study demonstrates that a reliable bone-cancer predictor can be achieved by aligning and fusing complimentary morphology and molecular context with uncertainty awareness. In held-out tests, the combined system obtained Accuracy 0.926, Macro-F1 0.914, AUROC 0.965, AUPRC 0.958, Brier 0.067, and ECE 0.018. A robust baseline is already deployable with radiograph-only performance (AUROC 0.940) in situations where only X-rays are available; incorporating WSIs and omics helps to fill in any remaining gaps and enhances calibration. Performance falls smoothly under corruption, stain jitter, and batch shift. On external sites, the fused AUROC stayed close to 0.958, indicating generalizability. Results from ablations demonstrated that reliability weighting, co-attention, and MMD + InfoNCE alignment are critical to these improvements. By improving clinical utility at high FN cost, a calibrated operating point θ* achieved

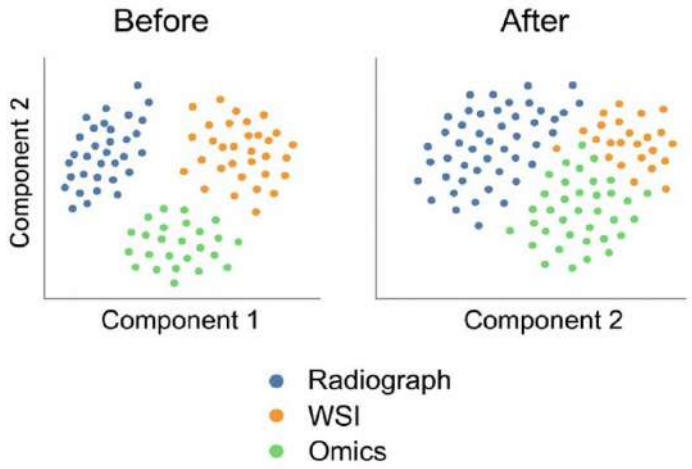

Figure 5: Latent Alignment Visualization: UMAP/t-SNE of radiograph/WSI/omics embeddings before vs after MMD+InfoNCE.

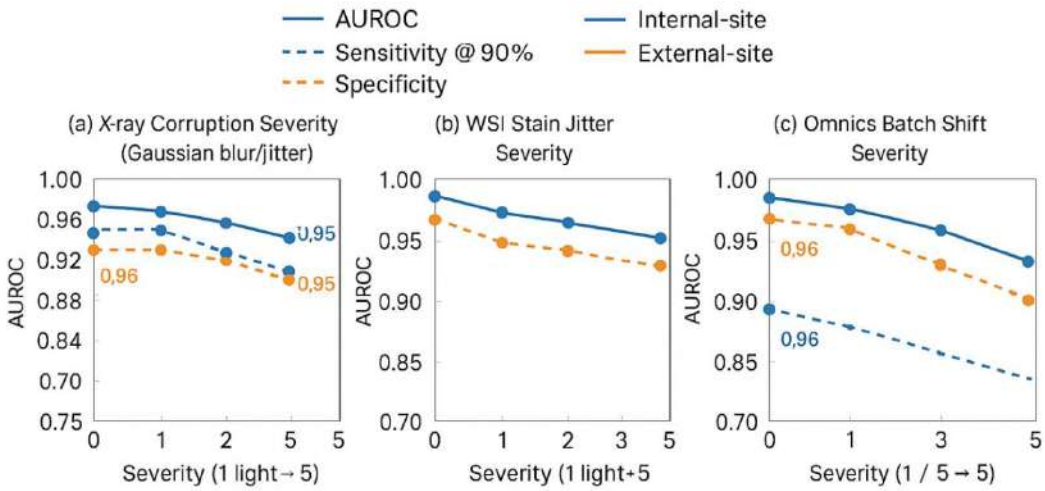

Figure 6: Robustness Curves: AUROC (and/or Sensitivity at fixed Specificity) vs corruption/stain/batch severity; external-site overlay.

94.1% Coverage, 0.936 Sensitivity, 0.903 Specificity, 0.907 PPV, and 0.933 NPV.

Imperfect modalities coupling, possible site bias in pathology and omics cohorts, and the lack of prospective, multi-center evaluation are some of the limitations. To utilized practical reliability scores, but more ethical Bayesian methods could enhance uncertainty estimations even further. While Grad-CAM, tile attention, and gene/pathway attributions have enhanced interpretability, systematic human-factor studies are still necessary. The evaluated datasets primarily represent osteosarcoma. Therefore, the model's generalizability to other bone malignancies, such as Ewing sarcoma or chondrosarcoma, has not been fully established. Further validation on more diverse subtype-specific cohorts is needed.

FUTURE SCOPE (FEATURE ROADMAP)

1. Prospective, multi-center trials with paired radiograph–WSI–omics per patient and standardized protocols.
2. Self-supervised pretraining on large radiology and histology corpora to improve low-label regimes.
3. Robust domain adaptation, e.g., stain/physics-aware augmentations and stronger batch correction for omics.
4. Federated learning to respect data governance while scaling across hospitals.
5. Active learning loops to prioritize pathology review for the most informative slides or tiles.
6. Integrated reporting: calibrated risk, human-readable rationales, and suggested next steps (e.g., biopsy).
7. Extended endpoints beyond detection—subtype prediction, response forecasting, event-free survival—turning the platform into a longitudinal decision-support engine.

By centering reliability, alignment, and graceful modality handling, the proposed framework offers a practical path from single-view screening to holistic, multimodal bone-cancer intelligence.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study utilized publicly available, de-identified datasets. No direct human or animal subjects were involved. Therefore, formal ethical approval and informed consent were not required.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The datasets used in this study are publicly available from Kaggle (Bone Cancer Detection dataset), The Cancer Imaging Archive (TCIA), and the TARGET-OS repository. All data generated or analyzed during this study are included in this published article.

The supplementary table can be downloaded from the journal website along with the article.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHORS' CONTRIBUTIONS

Dr. Shanker M.C. contributed to conceptualization, methodology, and manuscript drafting. Dr. V. Gokula Krishnan contributed to study design, validation, supervision, and manuscript review. Dr. D. Arul Kumar contributed to data analysis and technical validation. M. Bhuvaneswari contributed to implementation and experimental evaluation. Dr. K. Sathyamoorthy contributed to results interpretation and manuscript editing. Dr. B. Prathusha Laxmi contributed to literature review and final manuscript proofreading. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

The authors acknowledge the contributors of the Kaggle Bone Cancer Detection dataset, The Cancer Imaging Archive (TCIA), and the TARGET initiative for providing open-access data resources that enabled this research.

REFERENCES

[1] Sampath K, Rajagopal S, Chintanpalli A. A comparative analysis of CNN-based deep learning architectures for early diagnosis of bone cancer using CT images. Scientific Reports 2024; 14(1): 2144. https://doi.org/10.1038/s41598-024-52719-8

[2] Aarthy R, Muthupriya V, Balaji GN. Detection of bone cancer based on a four-phase framework generative deep belief neural network in deep learning. Alexandria Engineering Journal 2024; 109: 394-407. https://doi.org/10.1016/j.aej.2024.08.094

[3] Shrivastava A, Nag MK. Enhancing bone cancer diagnosis through image extraction and machine learning: a state-of-the-art approach. Surgical Innovation 2024; 31(1): 58-70. https://doi.org/10.1177/15533506231220968

[4] Murugan S, Patil SS, Reddy RK, KN SK. A recent survey on bone cancer detection using deep learning techniques. In: Proc. of the Second International Conference on Advances in Information Technology (ICAIT). IEEE 2024; pp. 1-6. https://doi.org/10.1109/ICAIT61638.2024.10690575

[5] Gunanithi S, Ilavarasan S, Karthika RN. Revolutionizing osteosarcoma diagnosis: a comparative analysis of deep learning models for precise bone cancer detection using multi-modal medical imaging. In: Int. Conf. Summit on Artificial Intelligence. Singapore: Springer 2024; pp. 197-207. https://doi.org/10.1007/978-981-97-7592-7_16

[6] Deng S, Huang Y, Li C, Qian J, Wang X. Auxiliary diagnosis of primary bone tumors based on machine learning model. Journal of Bone Oncology 2024; 49: 100648. https://doi.org/10.1016/j.jbo.2024.100648

[7] Zhao X, Dong YH, Xu LY, Shen YY, Qin G, Zhang ZB. Deep bone oncology diagnostics: computed tomography based machine learning for detection of bone tumors from breast cancer metastasis. Journal of Bone Oncology 2024; 48: 100638. https://doi.org/10.1016/j.jbo.2024.100638

[8] Baswaraju S, Thirumalraj A, Manjunatha B. Unlocking the potential of deep learning in knee bone cancer diagnosis using MSCSA-Net segmentation and MLGC-LTNet classification. In: Sustainable Development Using Private AI. CRC Press 2024; pp. 190-213. https://doi.org/10.1201/9781032716749-10

[9] Singh J, Patel P, Ingole BS, Inaganti R, Ramineni V, Krishnappa MS, *et al*. Advanced computational methods for pelvic bone cancer detection: efficacy comparison of convolutional neural networks. In: Proc. of IEEE 17th Int. Symp. on Embedded Multicore/Many-core Systems-on-Chip (MCSoC). IEEE 2024; pp. 287-293. https://doi.org/10.1109/MCSoC64144.2024.00055

[10] Srividya K, Reddy GV, Bakki V, Adilakshmi T. AI-based bone cancer detection using image processing and CNN. In: Int. Conf. on Cognitive Computing and Cyber Physical Systems. Cham: Springer Nature Switzerland 2024; pp. 283-304. https://doi.org/10.1007/978-3-031-77078-4_23

[11] Wang H, He Y, Wan L, Li C, Li Z, Li Z, *et al*. Deep learning models in classifying primary bone tumors and bone infections based on radiographs. NPJ Precision Oncology 2025; 9(1): 72. https://doi.org/10.1038/s41698-025-00855-3

[12] Xie W, Wang X, Liu M, Mai L, Shangguan H, Pan X, *et al*. A novel two-step classification approach for differentiating bone metastases from benign bone lesions in SPECT/CT imaging. Academic Radiology 2025. https://doi.org/10.1016/j.acra.2025.06.010

[13] Curto-Vilalta A, Schlossmacher B, Valle C, Gersing A, Neumann J, von Eisenhart-Rothe R, *et al*. Semi-supervised label generation for 3D multi-modal MRI bone tumor segmentation. Journal of Imaging Informatics in Medicine 2025. https://doi.org/10.1007/s10278-025-01448-z

[14] Wang K, Han Y, Ye Y, Chen Y, Zhu D, Huang Y, *et al*. Mixed reality infrastructure based on deep learning medical image segmentation and 3D visualization for bone tumors using DCU-Net. Journal of Bone Oncology 2025; 50: 100654 https://doi.org/10.1016/j.jbo.2024.100654

[15] Li B, Xu D, Lin H, Wu R, Wu S, Shao J, *et al*. Domain adaptive detection framework for multi-center bone tumor detection on radiographs. Computerized Medical Imaging and Graphics 2025; 123: 102522. https://doi.org/10.1016/j.compmedimag.2025.102522

[16] Kaggle. Bone cancer detection dataset. Available from: https://www.kaggle.com/datasets/ziya07/bone-cancer-detection-dataset

Received on 10-12-2025 Accepted on 08-01-2026 Published on 30-01-2026

https://doi.org/10.30683/1929-2279.2026.15.05