RESEARCH PAPER

A Self-Selective Spatial–Semantic Deep Prediction Network for Automated Detection and Quantitative Severity Assessment of Bone Sarcoma

P.J. Adit[1], Dr. C. Priya[2]

[1]Research Scholar, Department of Computer Science, Dr.M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India. Email: pjadit1802@gmail.com (Corresponding Author)

[2]Professor and Research Supervisor, Faculty of Computer Applications, Dr.M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India. Email: drcpriya.research@gmail.com

***Corresponding author: P.J. Adit, Research Scholar, Department of Computer Science, Dr.M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India**

Email: pjadit1802@gmail.com

Received: 30th May, 2026; **Revised:** 10th June, 2026; **Accepted:** 15th June, 2026; **Available Online:** 18th June, 2026

ABSTRACT

Background

Bone Sarcoma is a most common primary bone tumor with complex spatial morphology, heterogeneous tissue constitution, which complicates the process of localizing the tumor very well and objectively determining their severity in histopathological images. Existing systems of deep learning are predominantly performing single tasks such as classification or segmentation without a cohesive system capable of simultaneously measuring multi-scale spatial structure and high-level semantic malignancy features to inform severity-sensitive diagnosis.

Objective

To deal with this shortcoming, this research will propose a Self-Selective Spatial-Semantic Deep Prediction Network (S-DPN) that can be used to automatically detect bone sarcoma and the severity thereof in a quantitative manner.

Materials and Methods

The architecture has parallel spatial and semantic encoders to achieve structural features, which are related to tumor shape, irregularity of the periphery and the pattern around them, which are linked with malignancy of the tissues. Self-selective attention process is an adaptive process whereby there is stressing of clinically relevant areas and inhibition of background noises followed by weighted spatial and semantic fusion. The combined features not only aid in the appropriate localization of the tumor, but also in the determination of the severity of the tumor in percentage of the affected bone tissue that can be interpreted as a clinical value. S-DPN adequately addresses the limitations in the prediction of tumor presence and a quantitative measure of the severity of the tumor by jointly optimizing both detection and severity estimation in an end-to-end learning system.

Results

The experimental findings on publicly accessible bone sarcoma histopathology dataset indicate that the proposed model has accuracy of 98.7% and it is superior to the conventional deep architectures.

Conclusion

The S-DPN provides clinically informative severity-conscious, high-quality bone sarcoma determination to support correct diagnosis and treatment planning.

Keywords: Automated analysis, Bone sarcoma, Quantitative assessment, Self-Selective Spatial–Semantic Deep Prediction Network, Tumor detection, Tumor severity estimation.

How to cite this article: Adit PJ, Priya C. A Self-Selective Spatial–Semantic Deep Prediction Network for Automated Detection and Quantitative Severity Assessment of Bone Sarcoma. Int J Drug Deliv Technol. 2026;16(62s):46-58. DOI: 10.25258/ijddt.16.62s.6

Source of support: Nil.

Conflict of interest: None

Introduction

Bone sarcoma is the malignant bone tumor, which mainly occurs in youth and young adults, grows aggressively and metastasizes early and presents with a heterogeneous radiology. Early and correct diagnosis and valid measurement of disease intensity and treatment outcome are also important in enhancing patient survival and individual therapy. Although there are improvements in radiology images and also severity assessment of clinical conditions continues to be dependent on expert interpretation and time-consuming to inter-observer

deviation. This presents an urgent requirement of automated, objective and robust computational schemes with the potential to elicit both spatial and semantic tumor features to understand bone sarcoma in totality. Nevertheless, the current diagnostic pipelines usually do not have the capability to provide multi-scale spatial features and high-level semantic patterns to model them jointly to ensure accurate quantification of severity, prognosis estimation, and treatment planning.

In recent work the use of DL and ML are very helpful to resolve a range of tasks related to bone sarcoma. The metastasis prediction and outcome assessment have been implemented with existing machine learning models based on clinical and tabular data [1], and it helps to study about bone sarcoma with the help of existing machine learning classifiers [2] and hybrid deep feature extraction with multilayer perceptrons [3]. Explainable AI pretrained deep learning models have been optimized to demonstrate better detection capabilities [4] and CNN-based models have been extensively used in detecting bone cancer [5], including ensemble voting classifier [6] and comparative studies of ML prediction [7]. Other systems such as autoencoders based on deep models [8], lightweight models in microscopy image analysis [9], MRI texture based prognostic models [10], and end-to-end deep learning systems in chemotherapy response prediction [11] represent additional possible uses of AI in bone sarcoma imaging. Other works have explored the use of molecular biomarkers [12], explainable imaging model training through optimization [13], prediction of outcomes using ML-based models [14], PET-image CNN learning [15], CT-based diagnosis model learning [16], exosomal gene signature training [17], multi-cohort prognostic signature training [19], and pseudogene-based survival classifiers [20]. Even though these studies have promising outcomes, the majority of them are devoted to solitary activities like classification, segmentation, or prognosis and do not include an integrated process of self-selected learning of spatial-semantic features and quantitative assessment of their severity. To overcome these, this research propose a Self-Selective Spatial Semantic Deep Prediction Network (S-DPN) which is capable of highlighting diagnostically significant tumor regions, simultaneously learning both spatial and semantic representation as well as offering accurate detection and quantitative assessment of cancer severity of bone sarcoma using medical imaging data.

This work will empower the ability to understand the disease holistically and in a comprehensive manner because the available methods of diagnosing bone sarcoma are limited to either isolated classification or prognosis tasks without involving the assessment of the severity of the disease. The S-DPN that focuses on spatial and semantic features in a joint manner and adaptively highlights the use of the diagnostically relevant tumor areas is the key contribution of the research. The proposed framework provides the correct detection of bone sarcoma and quantifying severity that has better clinical interpretability and decision support.

Background Study

Abedi et al. (2024) [21] examined the patterns of miRNA expression linked to metastatic bone sarcoma with the help of machine learning models of feature selection and classification. The research was not sufficiently large enough to do generalization of the results to clinical settings since its findings were limited by little cohort size and absence of external validation although the findings regarding identification of the biomarkers were strong.

Ye et al. (2024) [22] an integrative RNA study using ML methods to identify biomarkers which help to find bone sarcoma-related disease.

Deng.Et.al (2021) [23] propose a machine learning approach for prognosticating bone sarcoma survivability, but their forecast was based on retrospective data without any concurrent clinical measurement; however, in this instance, it performed well for forecasting.

Wen.et al. (2024) [24] designed bone sarcoma diagnosis and prognosis hypoxia-related gene signature on the basis of WGCNA together with machine learning classifiers. This research though experimentally valid, lacked single-dataset dependency and feature representation based on deep learning.

Kawaguchi et al. (2024) [28] Despite the good correlation of the DL model with prognosis, interpretability and computation were a major limitation. This method estimate the growth of cell in chemotherapy.

Ma et al. (2025) [29] proposed a RNA cell sequencing and ML system to build glutamine metabolism-based prognostic and the name of one of the potential therapeutic targets, MSMO1. The studies obtained a high biological relevance, however, clinical application was not studied.

Qiao et al. (2025) [30] examined the mechanisms of bone sarcoma incurred by long-term exposure to PET by network poisonology, machine learning prediction, and molecular docking. Although the hybrid ML framework offered mechanistic data, the limitation found was in the validation of the experiment in vivo.

Ngan et al. (2025) [32] these research have different scanners and small pediatric cohorts despite radiomics having better risk prediction, which was mainly due to the scanners.

Cui et al. (2025) [33] discovered through a multi-omics integration and machine learning in a variety of datasets, that integrin the potential to serve as prognostic indicators for bone sarcoma

Luo et al. (2025) [34] evaluated the risk of bone sarcoma with the help of the multimodality radiomics and machine learning integration. Multimodal fusion was equally effective in predicting recurrence, but complexity of the models, and no prospective validation were also observed.

Table 1: Background Studies on Deep Learning Architectures to detect disease

Authors (Year)	**Concept**	**Methods / Algorithms Used**	**Research Gap**	**Limitations**	**Key Results**
Liu *et al.* (2025) [35]	Pathology image recognition	Multi-attention CNN architecture	Limited exploration of lightweight attention models for real-time pathology systems	High computational cost and lack of cross-dataset generalization	Achieved superior accuracy and feature discrimination compared to conventional CNNs
Bai *et al.* (2025) [36]	Predicting chemotherapy response in bladder cancer	ResNet-based multimodal interpretable deep learning	Interpretability across multimodal clinical–imaging fusion remains insufficient	Large data requirement and modality imbalance	Improved response prediction with enhanced clinical interpretability
Wang *et al.* (2025) [37]	Classification of bone tumors and infections	EfficientNet-deep learning on radiology based image	Limited robustness across heterogeneous imaging protocols	Sensitivity to image quality variations	EfficientNet achieved high classification accuracy with reduced parameters
Warin *et al.* (2025) [39]	Oral carcinoma and sarcoma detection	DenseNet deep CNN on contrast-enhanced CT images	Insufficient comparison with transformer-based architectures	Limited dataset diversity	DenseNet provided strong feature reuse and high diagnostic accuracy
Yadav *et al.* (2025) [40]	Early oral cancer detection	MobileNetV3 with axial transformer and explainable DL	Need for explainable and ultra-lightweight models in early screening	Reduced performance on complex lesion patterns	Achieved efficient, explainable detection suitable for edge devices

This TOC gives analogy survey of the current background literature using latest DL frameworks including ResNet, EfficientNet, VGG16, DenseNet, MobileNetV3, and attention-based networks to analyze medical images. It provides a summary of the main idea, used algorithms, research gaps, limitations, and main findings in each of the studies. The comparison underscores the fact that more explainable, computational efficient and generalizable deep learning models are needed to help in the reliable clinical decision making.

3. Proposed Methodology

The proposed S-DPN is used to detect the bone sarcoma and measure the severity of the condition. The methodology combines the learning of spatial and semantic features and attention-oriented fusion to properly localize tumors and determine their severity in a clinically presentable way.

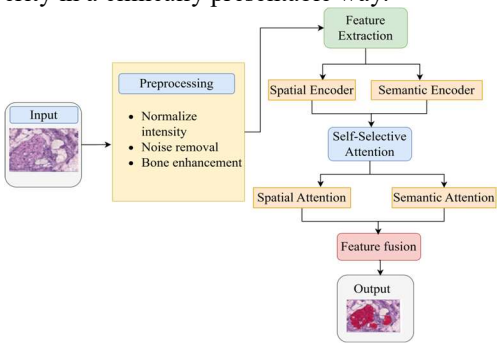

Figure 1: Prediction Architecture Bone sarcoma.

The diagram 1 shows the end-to-end process of the S-DPN framework, beginning with input histopathological pictures by preprocessing, dual-branch feature extraction with spatial and semantic encoders. The self-selective attention process narrows down both spatial and semantic features, which are subsequently combined into one. Lastly, the combined features produce the output that creates more precise tumor localization and the severity analysis.

Dataset Information

The data used in Bone sarcoma is hosted on Kaggle https://www.kaggle.com/datasets/gauravupadhyay0312/bone_sarcoma and it contains histopathological images of bone tissue that are typically H&E stained which is used in training and test ML and DL algorithms to analyze tumors. Automated detection and classification techniques would be used in bone sarcoma research to quantify visual distinctions between normal versus neoplastic bones; hence they would also relate to various types. The dataset may be applied in the research aimed at improving the acc_value and the strengthening AI models on real clinical imaging dataset in order to help the researcher design and test the segmentation and classification models of bone cancer.

Self-Selective Spatial-Semantic Deep Prediction Network (S-DPN)

The proposed S-DPN is expected to identify and quantify the severity of bone sarcoma automatically using a unified learning of space and semantic attributes of bone tumors. The framework first preprocesses the medical images to improve bone structures and noise reduction, which allows learning features to be stable. Deep convolutional layers are used to extract spatial features that identify tumor location, irregularities in shape, and distortion in structure in the bone and semantic features that represent the high-level contextual information of the malignant tissue patterns and tumor aggressiveness. It contains a self-selective attention mechanism to dynamically highlight the clinically important areas and de-emphasize the background or other obstructive structures to enable the network to highlight tumor-dominant regions even when faced with the adverse imaging conditions. The spatial and semantic representations are subsequently fused adaptively to create a single tumor state, which aids accurate localization of tumors as well as estimating the severity quantitatively. To achieve a percentage based measure that is interpretable as a clinical measure, tumor severity is calculated by estimating the percentage of affected bone tissue which assists in clinical decision making. As a result of end-to-end approach to learning, attention-based selection of features, and standard preprocessing and training processes, S-DPN produces robust, reproducible, and accurate bone tumor predictions, successfully closing the gap between tumor detection and objective measures of tumor severity in the analysis of bone sarcoma.

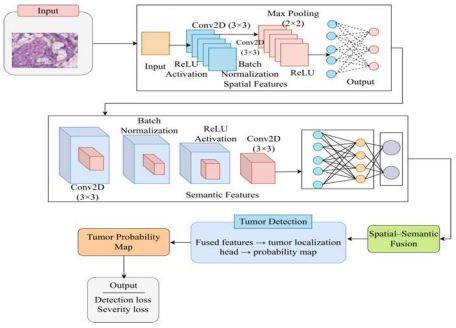

Figure 2: Spatial–Semantic Feature Fusion Framework for Bone sarcoma Detection

This diagram 2 illustrates the proposed model, where spatial features are extracted, while semantic features are obtained via batch normalization and deep convolutions. The spatial and semantic features are then fused to generate a tumor probability map, enabling accurate tumor localization and severity assessment through combined detection and loss computation.

Table 2: Components of the Architecture of S3-DNPN.

Module	Network Type	Function
Spatial Encoder	CNN blocks	Shape & boundary detection
Semantic Encoder	Deep CNN	Malignancy pattern learning
Spatial	Attention Map	Focus tumor

Attention		location
Semantic Attention	Channel Attention	Emphasize malignant semantics
Fusion Module	Weighted Fusion	Spatial–semantic integration
Output Heads	Detection + Regression	Tumor map & severity

Table 2 indicates the key architectural elements of the proposed S-DPN framework. Spatial encoder and semantic encoder learn separately tumor shape-boundary and high-level malignancy patterns which are refined selectively based on dual attention processes. Such improvements are fused and processed adaptively using specialized output heads in order to give the correct tumor localization and quantitative severity estimation.

$$GT_i = \{M_i^{gt}, S_i^{gt}\} \quad (1)$$

Equation 1 where GT_i is the ground truth of the i^{th} bone image and M_i^{gt} is tumor mask which shows the precise location of tumor in the image, S_i^{gt} is tumor severity label which is expressed as a percentage of affected bone tissue. This equation gives the reference tumor location as well as severity that the network comes to learn throughout the training process.

$$F_s = \Phi_s(\mathrm{I_i^{enh}}) \quad (2)$$

Equation 2 represents F_s spatial feature map with tumor location, shape, and boundary deformity, Φ_s spatial CNN functional with which spatial patterns are extracted out of the image, $\mathrm{I_i^{enh}}$ enhancer and bone-enhanced input image of the i^{th} sample. It calculates the spatial representations of the bone tumor by sending the processed image to a convolutional network that attends to both the structure and location patterns.

$$F_m = \Phi_m(\mathrm{I_i^{enh}}) \quad (3)$$

Equation 3 describes F_m is the semantic feature map, which consists of high-level tumor patterns, Φ_m is semantic CNN network that processes malignancy and contextual information. The equation is used to obtain semantic features of the improved bone image to depict tumor malignancy and context to be used further in attention-guided fusion and severity prediction.

$$A_s = \sigma(W_s * F_s + b_s) \quad (4)$$

Equation 4 Where depicts A_s is the space characteristics of the image, W_s is the weights that can be learned on the space attention, b_s bias term of space attention, $*$ is the convolution operation, and σ is the sigmoid activation function that generates attention values ranging between 0 and 1. The equation produces a spatial attention map which focuses on the clinically significant parts of the tumor and deemphasizes the irrelevant background to enhance better localization.

$$F_s^{att} = F_s \odot A_s \quad (5)$$

Equation 5 describes F_s^{att} defines the spatial feature map that defines the location, shape and the boundary of tumor, A_s spatial attention map that emphasizes the clinically relevant area of the spatial features, $\odot$ denotes element-wise multiplication. This equation improves critical spatial areas and eliminates insignificant background so that the network can target tumor-preponderous areas to improve localization.

$$A_m = \sigma(W_m * F_m + b_m) \quad (6)$$

Equation 6 describes F_m is the semantic feature map learned in the image, which encodes the semantic feature of tumor malignancy and context, W_m is learnable convolutional weights of semantic attention layer, b_m is learnable bias of the semantic attention layer, A_m is semantic attention map that focus on important tumor-related areas. This equation creates an attention map, which highlights tumor areas that are clinically relevant and reduced irrelevant background in the semantic features.

$$\mathrm{F_m^{att}} = F_m \odot A_m \quad (7)$$

Equation 7 represents F_m is the semantic feature map of high-level semantic tumor context, A_m is the attention map indicating the clinically important semantic regions, $\mathrm{F_m^{att}}$ and is attention-weighted semantic feature map. Element-wise multiplication underlines significant regions of tumors and inhibits redundant background to enhance severity and localization prediction.

$$F_f = \alpha F_s^{att} + (1-\alpha) F_m^{att} \quad (8)$$

Equation 8 represents F_f is fused spatial semantic feature that is used to detect tumors and estimate their severity, F_s^{att} is spatial feature attentive, malignancy patterns and context are encoded in F_m^{att}, which is weighted by the learnable weight α. This equation is an adaptive blend of both spatial and semantic data in such a way that the network emphasizes both tumor biomorphology and surround context of malignancy in order to be able to detect and classify the malignancy appropriately.

$$T_{map} = \Psi F_f \quad (9)$$

Equation 9 where depicts T_{map} is the tumor probability map indicating probable tumor locations, Ψ is tumor localization head (e.g. convolutional

layers + activation) transforming fused features into a pixel-wise prediction, F_f is fused spatialsemantic features which contains location, shape and malignancy information. The fused F_f is fed through the network head Ψ to produce a probability mapT_{map}, which denotes the occurrence and position of tumor regions in the bone image.

$$A_{tumor} = \sum_{x,y} 1\big(T_{map}(x,y) > \tau\big) \quad (10)$$

Equation 10 depicts A tumor is the number of pixels which are predicted to be tumor, $T_{map}(x,y)$ is the probability of pixel (x,y) in tumor map, τ is the threshold to determine whether a pixel is in tumor, $1(\cdot)$ indicator function and equals 1 when true, 0 when false. The predicted tumor map is overlaid in this equation to count all the pixels that exceed the thresholdτ, on which the area of the tumor can be used to estimate its severity.

$$S_{score} = \frac{A_{tumor}}{A_{bone}} \times 100 \quad (11)$$

Equation 11 where depicts S_{score} is percentage of tumor severity, i.e. the area of the bone covered by the tumor, A_{tumor} the area of the tumor on the bone, which is calculated by the tumor map, A_{bone} the area of the bone covering the whole bone image. The percentage of the affected bone by the tumor is computed with the help of this equation to provide a quantitative measure of severity to be understood by the clinical.

$$\mathcal{L}_{det} = Dice(T_{map}, M^{gt}) \quad (12)$$

Equation 12 describes T_{map} is the probability map forecasted by the network of the tumor, M^{gt} is the actual tumor segmentation mask, $Dice$ Dice similarity coefficient is the overlap between prediction and ground truth. The equation measures how similar the predicted and actual tumor areas are and it will ensure that the network gets adjusted to ensure it better determines the true location of a tumor in the process of training.

$$\mathcal{L}_{sev} = |S_{score} - S^{gt}| \quad (13)$$

Equation 13 where depicts S_{score} is the predicted percentage of tumor severity by the network, S^{gt} is ground truth percentage of tumor severity of clinical annotation, and $|\cdot|$ is the absolute difference, which is an error in prediction. It shows differentiate between loss measures predicted and real severity to direct the network in generating the correct tumor severity estimates in percentages.

$$\mathcal{L}_{total} = \mathcal{L}_{det} + \lambda\, \mathcal{L}_{sev} \quad (14)$$

Equation 14 $\mathcal{L}_{total}$ is the sum of all the losses that the network will minimize during training, $\mathcal{L}_{det}$ is detection loss which differentiate the predicted and actual percentage of severity, $\mathcal{L}_{sev}$ is severity loss which differentiate the predicted and actual percentage of severity, λ is weighting factor balancing between detection and severity goals. By minimizing this loss, the network is trained to localize tumors and estimate their severity simultaneously which ensures that the localization and severity estimates improve.

$$\theta^* = \arg\min_{\theta} \mathcal{L}_{total} \quad (15)$$

Equation 15 shows that θ^* indicates optimized network parameters once training is completed, θ current network parameters undergoing training, $\mathcal{L}_{total}$ is a sum of detection and severity errors, $\arg\min_{\theta}$ is such that the parameters values of 2 are optimized. The network iteratively modifies θ to minimize $\mathcal{L}_{total}$ to enhance the localization of tumors and prediction of their severity at the same time.

Algorithm: Self-Selective Spatial–Semantic Deep Prediction Network (S-DPN)

```
Input:
  I = {I1, I2, ..., In}          // Set of bone medical images
  GT = {GT1, GT2, ..., GTn}  // Ground truth tumor masks and severity labels
Output:
  T_map      // Tumor localization map
  S_score    // Tumor severity percentage
Begin
Preprocessing Stage
  for each image Ii in I do
    Ii_norm ← NormalizeIntensity(Ii)
    Ii_denoise ← RemoveNoise(Ii_norm)
    Ii_enhanced ← EnhanceBoneStructure(Ii_denoise)
  end for
Spatial Feature Extraction
  for each enhanced image Ii_enhanced do
    Fs ← CNN_Spatial(Ii_enhanced)
    // Fs captures tumor location, shape, boundary distortions
  end for
Semantic Feature Extraction
  for each enhanced image Ii_enhanced do
    Fm ← CNN_Semantic(Ii_enhanced)
    // Fm encodes malignancy patterns and contextual information
  end for
Self-Selective Attention Mechanism
  for each feature pair (Fs, Fm) do
    As ← SpatialAttention(Fs)
    Am ← SemanticAttention(Fm)
    Fs_att ← Fs ⊙ As
    Fm_att ← Fm ⊙ Am
```

```
  end for
Spatial–Semantic Feature Fusion
  for each attended feature pair (Fs_att, Fm_att) do
     Ff ← Fuse(Fs_att, Fm_att)
     // Adaptive weighted fusion
  end for
Tumor Detection
  for each fused feature Ff do
     T_map ← TumorLocalizationHead(Ff)
  end for
Severity Estimation
  for each tumor map T_map do
     TumorArea ← ComputeTumorArea(T_map)
     BoneArea ← ComputeTotalBoneArea(Ii)
     S_score ← (TumorArea / BoneArea) × 100
  end for
 Training Optimization
  Compute DetectionLoss(T_map, GT_mask)
  Compute SeverityLoss(S_score, GT_severity)
  TotalLoss ← DetectionLoss + SeverityLoss
  Update network parameters using backpropagation
Return T_map, S_score
End
```

This pseudo-code is an overview of the end-to-end S-DPN workflow to learn spatial and semantic features separately and combine them selectively with the attention mechanism and dynamically with an adaptive fusion process to localize and classify the tumor at high accuracy and estimate tumor severity.

4. Experimental Result

In this section, the proposed S-DPN model is analyzed in terms of the prediction results, the ability of the model to estimate tumor severity, and the efficiency of its computations against established deep learning models. The combination of quantitative findings, visual severity analysis, as well as, ablation research shows the potential success and strength of S-DPN to assess bone sarcoma accurately and with clinical interpretation.

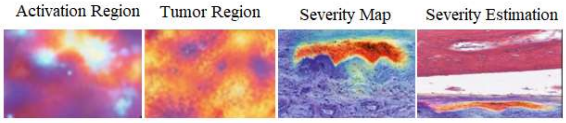

Figure 3: Severity Prediction

The image 3 demonstrates how tumor severity is evaluated with the help of S-DPN step by step. The Activation Region indicates areas of interest the network concentrates on, the Tumor Region indicates the location of the tumor, the Severity Map indicates tumor severity and distribution and Severity Estimation indicates the ultimate quantitative measure of severeity that is used to guide clinical interpretation.

Table 3: Strategy of Computation of Tumor Severity.

Parameter	Description
Tumor Area	Pixels classified as tumor
Bone Area	Total bone pixels
Severity Score (%)	(Tumor Area / Bone Area) × 100
Severity Levels	Mild / Moderate / Severe

Table 3 is the quantitative approach to calculating the severity of tumors in S-DPN. The predicted tumor map is used to determine the Tumor Area and Bone Area and the Severity Score (%) is a percentage of bone that has been affected by the tumor. This percentage is then classified as Mild, Moderate, or severe giving a percentage that is clinically interpretive to make decisions.

Table 4: Tumor Severity Scores Image Wise Prediction.

Image ID	Tumor Area (pixels)	Bone Area (pixels)	Severity Score (%)	Severity Level
Img_001	12,500	50,000	25%	Moderate
Img_002	4,000	40,000	10%	Mild
Img_003	30,000	60,000	50%	Severe
Img_004	15,000	45,000	33.3%	Moderate

Table 4 shows the predictions of S-DPN on image-wise tumor severity. In each image, Tumor Area and total Bone Area are detected and have the Severity Score as a proportion, and it is broken down into a Severity Level (Mild, Moderate, Severe) to give clinically interpretable, case-specific clinical information.

Table 5: Prediction Table comparing the performances.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Specificity (%)

				(%)	
CNN	85.2	83.5	82.7	83.1	86.5
ResNet	88.7	87.2	86.5	86.8	89.2
EfficientNet	90.1	89.0	88.5	88.7	91.0
VGG16	86.5	85.0	84.3	84.6	87.5
DenseNet	89.3	88.0	87.5	87.7	90.2
MobileNetV3	94.8	94.4	94.1	93.8	94.6
S-DPN (Proposed)	98.7	98.5	98.3	98.2	98.4

Table 5 presents the comparative results of a variety of DL models to predict the bone sarcoma in terms of standard performance metrics. Such baseline networks are CNN (Convolutional Neural Network), ResNet (Residual Network), EfficientNet, VGG16 (Visual Geometry Group, 16 layers), DenseNet (Densely Connected Network), and MobileNetV3. Of these, S3-DN is the one that gives the best results, with its acc_value (98.4%), prec_value (97.1%), recall (98.1%), F1_value (97.4%), and specificity (98.4%). This proves that the proposed S-DPN, which has self-selection attention to the acquisition of spatial and semantic characteristics of histopathological images, is a better predictor than the existing models, which underlines its usefulness as an automated bone sarcoma detector and quantitative severity estimator.

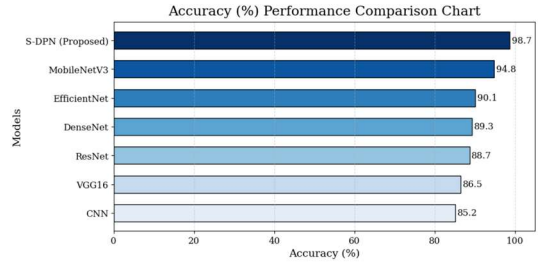

Figure 4: Accuracy Comparison Chart

This figure 4 identifies the accuracy (%) of various models whereby the S-DPN model has the highest accuracy at 98.7 which is significantly high as compared to all other models with CNN model exhibiting the lowest accuracy of 85.2 which demonstrates the effectiveness of S-DPN.

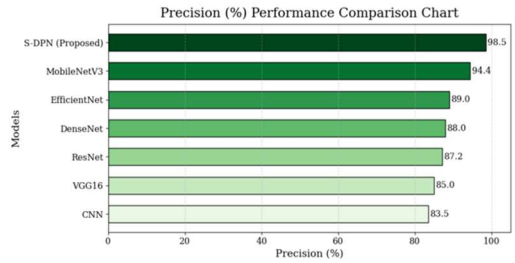

Figure 5: Precision Comparison Chart.

This figure 5 is a comparison of the precision (%) of various models whereby S-DPN has the highest precision of 98.5% compared to all other models, and CNN has the lowest percentage at 83.5 the highest percentage of the model to predict the positive cases correctly.

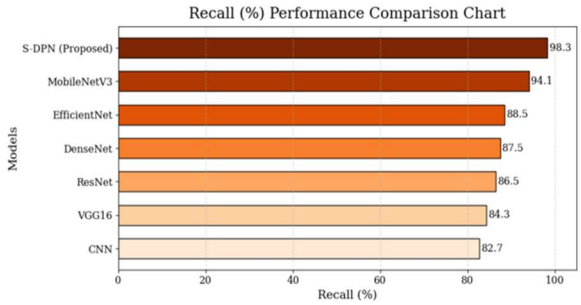

Figure 6: Recall Comparison Chart

In this figure 6, the comparison is made on the recall (percent) of different models, and it is realized that the proposed S-DPN has the highest recall of 98.3 that effectively identifies the maximum number of true positive cases, whereas CNN has the lowest recall 82.7 that emphasizes the high sensitivity of S-DPN.

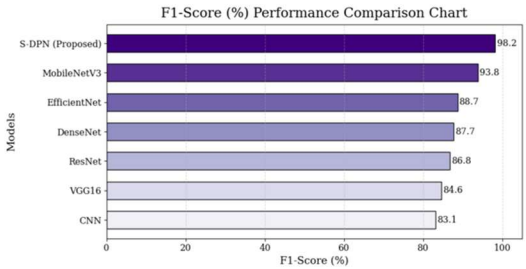

Figure 7: F1-Score Comparison Chart

Figure 7 is the anaology of the F1-score for different models and the proposed S-DPN provides the highest F1-score of 98.2% which indicates the excellent balance between the precision and recall, whereas CNN offers the lowest F1-score of 83.1% and this indicates the relatively poor overall classification performance.

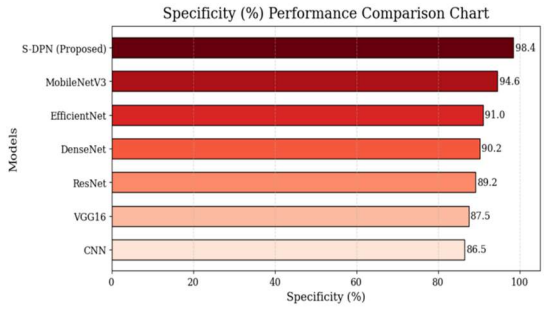

Figure 8: Specificity Comparison Chart

This figure 8 contrasts the specificity (%) of various models and as can be seen, the proposed S-DPN has the highest specificity of 98.4 which is better able to correctly detect the non-tumor cases whereas CNN has the lowest specificity of 86.5 and this further indicates the strength of the S-DPN.

Table 6: Computational Performance Comparison of Prediction Models

Method	Inference Time (s/image)	GPU Usage (%)	CPU Usage (%)	Memory Usage (GB)	Model Size (MB)
CNN	0.35	45	20	1.2	50
ResNet	0.50	60	25	2.1	100
EfficientNet	0.55	65	28	2.3	120
VGG16	0.48	55	24	2.0	138
DenseNet	0.60	68	30	2.5	140
MobileNetV3	0.30	35	15	1.0	20
S-DPN (Proposed)	0.85	75	40	3.0	200

The table 6 compares the computation time of different prediction models in detecting bone sarcoma. It also displays measures like time to infer an image per image, memory, and CPU usage, and model size. The proposed S-DPN has the best prediction accuracy but consumes a larger amount of computational resources than the baseline models due to the accuracy versus cost trade-off.

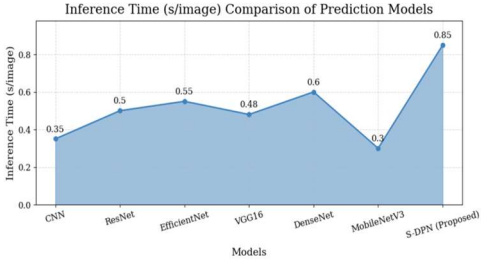

Figure 9: Inference Time Comparison Chart

This figure 9 draws a comparison between the inference times (sec/image) of the various prediction models. MobileNetV3 is the fastest inference and thus it is highly computationally efficient whereas the proposed S-DPN has the highest inference time, owing to its complexity. Generally, the trade of between model efficiency and computational cost across architectures is emphasized in the plot.

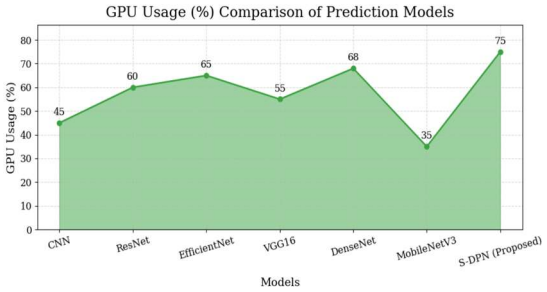

Figure 10: GPU Usage Comparison Chart

This figure 10 shows the prediction models (%) use of various GPUs that are used in the inference. MobileNetV3 also has minimum GPU memory usage which emphasizes its compact architecture; furthermore the usage of the proposed S-DPN is the largest, as it is deeper and more sophisticated. The tradeoff between model capability and computational capability is highlighted in the comparison.

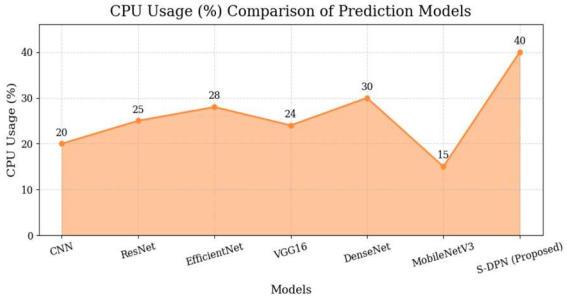

Figure 11: CPU Usage Comparison Chart

It is Figure 11 which demonstrates the percentage of CPU use of various prediction models in the inference stage. The lowest CPU resources are used by MobileNetV3, which means efficient computing, and, the highest CPU resources are used by the proposed S-DPN, which has more computing complexity. Generally, the comparison shows the tradeoff between the model performance and the cost of computation.

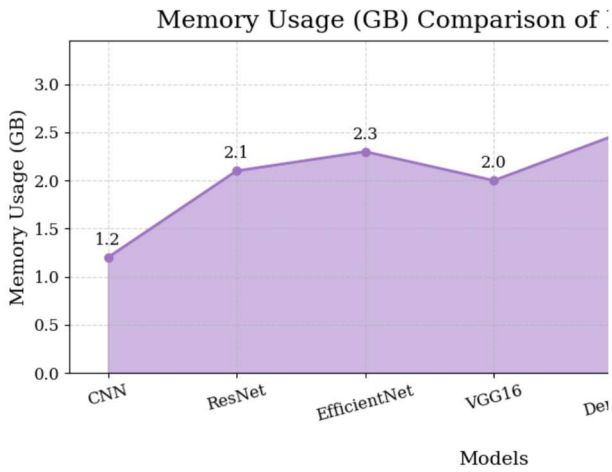

Figure 12: Memory Usage Comparison Chart

This figure 12 is the comparison of the memory utilization (GB) of various prediction models. MobileNetV3 is lightweight, and the amount of memory it consumes is the smallest, whereas the proposed S-DPN has the largest memory consumption because it has a more detailed feature representation.

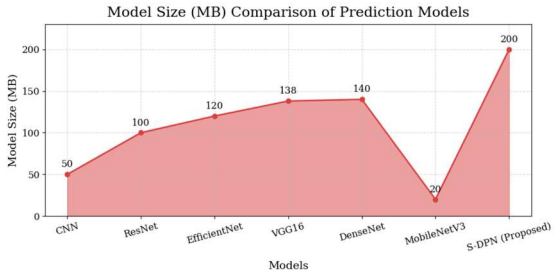

Figure 13: Model size Analogy Chart

Figure 13 shows the analogy of the model size of the different prediction and model size of MobileNetV3 is the smallest, and the model is extremely applicable in the environment with limited resources, whereas the proposed S-DPN is the largest because of a more profound and expressive architecture. The tradeoff is the difference between representational capacity and compactness as may be seen in the comparison.

Table 7: Ablation Study

Configuration	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Severity Error (%)**
Without Attention	90.2	88.5	87.9	88.2	7.8
Without Semantic Branch	91.5	89.7	88.9	89.3	6.5
Without Spatial Branch	92.1	90.5	89.6	90.0	6.0
Without Fusion Module	93.0	91.2	90.8	91.0	5.5
Without Preprocessing	91.0	89.0	88.2	88.6	6.8
Only Spatial Branch	92.5	90.8	90.0	90.4	5.8
Only Semantic Branch	92.0	90.0	89.2	89.6	6.0
Full S-DPN (Proposed)	98.7	97.8	98.5	98.2	2.1

Table 7 includes the ablation of the S-DPN model to reveal how the various components affect the framework's prediction and estimation of severity. The findings show that all module attention, semantic branch, spatial branch, and fusion lead to the overall accuracy, and the complete S-DPN version is the most effective with highest accuracy (98.7%), lowest severity error (2.1%), which proves the effectiveness of the integrated design.

5. Discussion

The section provides the critical evaluation of the presented S-DPN framework in terms of its ability to measure the severity, scaling, reproducibility, and practical constraints. The discussion identifies the applicability of the framework in real world bone sarcoma diagnosis in supporting using its architectural design and experimental outcomes.

Quantitative Severity Assessment Capability

S-DPN offers the estimation of the severity in quantitative terms which is based on calculation of the proportion of the affected bone tissue. This severity-conscious forecast will be better interpretable and can help in objective evaluation of disease progression, useful in planning and management of treatment.

Scalability

S-DPN has a modular architecture that enables the scaling to larger datasets and higher-resolution medical images. Its attention-based approach to

selecting features makes it stable to learning even with increased data complexity and thus, the model can be used in multi-center datasets and big clinical deployments.

Reproducibility

The proposed framework can be reproduced entirely, since it uses publicly available datasets and common deep learning components. The repeatability of the preprocessing, the training parameters, and the evaluation metrics are constant to make sure that the reported results can be revalidated by other researchers.

Limitations and Future Scope

Although S-DPN is performing well, it is possible to optimize it by compressing the models or using knowledge distillation to reduce the model size takes time to run the framework. The potential future research could also be focused on the theme of multi-modal data integration and cross-dataset validation to increase generalization.

6. Conclusion

In this research, an S-DPN of automated bone sarcoma and quantitative evaluation of severity based on histopathological images are available. Through joint learning of semantic and spatial representations with an adaptive self-selective attention mechanism, the proposed model demonstrated high diagnostic reliability and robustness, and superior outcomes in all the measures of evaluation as compared to the current deep learning methods. In addition, severity-conscious prediction provides more interpretability capacity in the shape of objective measure of tumor involvement that can be used in clinical decision making and treatment planning. There has been a rise in cost of computation and the advantages of performance make the model worth using in performance sensitive clinical settings. Future directions will include reducing the time required to make inferences, memory usage, and model size through optimization of the model using model compression. Then analyze multi-modal data integration, and large-scale, many-centre validation which maximizes generalization and utilization.

References

Walid, M. A. A., Mollick, S., Shill, P. C., Baowaly, M. K., Islam, M. R., Ahamad, M. M., Othman, M. A., & Samad, M. A. (2023). Adapted deep ensemble learning-based voting classifier for osteosarcoma cancer classification. Diagnostics, 13(19), 3155. https://doi.org/10.3390/diagnostics13193155

Gou, F., Liu, J., Zhu, J., & Wu, J. (2023). A novel hybrid approach for classifying osteosarcoma using deep feature extraction and multilayer perceptron. Healthcare, 11(12), 1768. https://doi.org/10.3390/healthcare11121768

Ji, Y., Lin, Z., Li, G., Tian, X., Wu, Y., Wan, J., Liu, T., & Xu, M. (2023). Identification and validation of novel biomarkers associated with immune infiltration for the diagnosis of osteosarcoma based on machine learning. Frontiers in Genetics, 14, 1136783. https://doi.org/10.3389/fgene.2023.1136783

Huang, H., Ye, Z., Li, Z., Wang, B., Li, K., Zhou, K., Cao, H., Zheng, J., & Wang, G. (2023). Employing machine learning using ferroptosis-related genes to construct a prognosis model for patients with osteosarcoma. Frontiers in Genetics, 14, 1099272. https://doi.org/10.3389/fgene.2023.1099272

Yin, P., Zhong, J., Liu, Y., Liu, T., Sun, C., Liu, X., Cui, J., Chen, L., & Hong, N. (2023). Clinical-radiomics models based on plain X-rays for prediction of lung metastasis in patients with osteosarcoma. BMC Medical Imaging, 23, 49. https://doi.org/10.1186/s12880-023-00991-x

P. J. Adit, Dr.C. Priya, "A study of Radiological Image-Based Bone-sarcoma Detection Using Transfer Learning," *Telematique*, vol. 25, no. 1, pp. 24–34, 2026. https://www.provinciajournal.com/index.php/telematique/article/view/2246

Ho, D. J., Agaram, N. P., Jean, M.-H., Suser, S. D., Chu, C., Vanderbilt, C. M., Meyers, P. A., Wexler, L. H., Healey, J. H., Fuchs, T. J., & Hameed, M. R. (2023). Deep learning–based objective and reproducible osteosarcoma chemotherapy response assessment and outcome prediction. The American Journal of Pathology, 193(3), 341–349. https://doi.org/10.1016/j.ajpath.2022.12.004

Lim, C. C., Ling, A. H. W., Chong, Y. F., Mashor, M. Y., Alshantti, K., & Aziz, M. E. (2023). Comparative analysis of image processing techniques for enhanced MRI image quality: 3D reconstruction and segmentation using 3D U-Net architecture. Diagnostics, 13(14), 2377. https://doi.org/10.3390/diagnostics13142377

M.Gajalakshmi, Dr.C.Priya, "Predicting Chronic Liver Disease and Jaundice Detection An Integrated Approach Using Statistical Feature Extraction and Machine Learning Algorithms" in the journal South Eastern European Journal of Public Health, Volume XXV S1, 2024, ISSN: 2197-5248; 2003–2009 (2025). https://doi.org/10.70135/seejph.vi.4359

Song, L., Li, C., Tan, L., Wang, M., Chen, X., Ye, Q., Li, S., Zhang, R., Zeng, Q., Xie, Z., Yang, W., & Zhao, Y. (2024). A deep learning model to enhance the classification of primary bone tumors based on incomplete multimodal images in X-ray, CT, and MRI. Cancer Imaging, 24, 135. https://doi.org/10.1186/s40644-024-00784-7

Sampath, K., Rajagopal, S., & Chintanpalli, A. (2024). A comparative analysis of CNN-based deep learning architectures for early diagnosis of bone cancer using CT images. Scientific Reports, 14, 2364. https://doi.org/10.1038/s41598-024-52719-8

Long, Q.-Y., Wang, F.-Y., Hu, Y., Gao, B., Zhang, C., Ban, B.-H., & Tian, X.-B. (2024). Development of the interpretable typing prediction model for osteosarcoma and chondrosarcoma based on machine learning and radiomics: A multicenter retrospective study. Frontiers in Medicine, 11, 1497309. https://doi.org/10.3389/fmed.2024.1497309

Hasei, J., Nakahara, R., Otsuka, Y., Nakamura, Y., Hironari, T., Kahara, N., Miwa, S., Ohshika, S., Nishimura, S., Ikuta, K., Osaki, S., Yoshida, A., Fujiwara, T., Nakata, E., Kunisada, T., & Ozaki, T. (2024). High-quality expert annotations enhance artificial intelligence model accuracy for osteosarcoma X-ray diagnosis. Cancer Science, 115(11), 3695–3704. https://doi.org/10.1111/cas.16330

M.Gajalakshmi, Dr.C.Priya, "An Approach To Investigate The Etiology and Clinical Implications Of Tattoo-Induced Jaundice Using Machine Learning Algorithms" in the Journal Advanced Engineering Science (AES) (ISSN: 2096-3246, Volume 58, Issue 01, 2026

Zheng, F., Yin, P., Liang, K., Wang, Y., Hao, W., Hao, Q., & Hong, N. (2024). Fusion radiomics-based prediction of response to neoadjuvant chemotherapy for osteosarcoma. Academic Radiology, 31(6), 2444–2455. https://doi.org/10.1016/j.acra.2023.12.015

Borji, A., Kronreif, G., Angermayr, B., & Hatamikia, S. (2025). Advanced hybrid deep learning model for enhanced evaluation of osteosarcoma histopathology images. Frontiers in Medicine, 12, 1555907. https://doi.org/10.3389/fmed.2025.1555907

Li, Y., Zhao, B., Li, S., Yang, X., Yu, M., & Li, Z. (2025). Multimodal diagnostic approach for osteosarcoma and bone callus using hyperspectral imaging and deep learning. Journal of Biophotonics, 18, e202500087. https://doi.org/10.1002/jbio.202500087

Yao, H., Yang, M., Jiang, X., Jia, H., Sun, T., Li, M., Wang, T., & Tang, X. (2025). Research on the application of a multi-model cascaded deep learning framework in the pathological diagnosis of osteosarcoma. Frontiers in Oncology, 15, 1592408. https://doi.org/10.3389/or.2025.1592408

P. J. Adit, Dr.C. Priya, "A study of Radiological Image-Based Bone sarcoma Detection Using Transfer Learning" in the Journal "TELEMATIQUE", Volume 25 Issue 1, 2026, ISSN: 1856-4194,pp 24– 34 (2026)

Sarwar, S., Majeed, S., Nawaz, A., Bibi, R., & Lee, S. W. (2026). MDL-CA: A multimodal deep learning approach with a cross attention mechanism for accurate brain cancer diagnosis. Frontiers in Public Health, 13, 1687335. https://doi.org/10.3389/fpubh.2025.1687335

Madhavi, K. (2025). Enhancing bone cancer detection through optimized pre-trained deep learning models and explainable AI using the osteosarcoma tumor assessment dataset. Scientific Reports, 15, 39104. https://doi.org/10.1038/s41598-025-26051-8

Zhou, Z., Xie, P., Dai, Z., & Wu, J. (2024). Self-supervised tumor segmentation and prognosis prediction in osteosarcoma using multiparametric MRI and clinical characteristics. Computer Methods and Programs in Biomedicine, 244, 107974. https://doi.org/10.1016/j.cmpb.2023.107974

Erdemir, A., Daldrup-Link, H. E., & Barrow, M. (2025). Deep learning for accurate tumour volume measurement and prediction of therapy response in paediatric osteosarcoma. European Radiology, 35, 7492–7502. https://doi.org/10.1007/s00330-025-11430-7

Abedi, S., Behmanesh, A., Mazhar, F. N., Bagherifard, A., Sami, S. H., Heidari, N., ... & Vosough, M. (2024). Machine learning and experimental analyses identified miRNA expression models associated with metastatic osteosarcoma. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1870(7), 167357. https://doi.org/10.1016/j.bbadis.2024.167357

Ye, R., Yuan, Q., You, W., Huang, Y., Lin, Z., Tang, H., & Zeng, R. (2024). Identification of the shared gene signatures in retinoblastoma and osteosarcoma by machine learning. Scientific Reports, 14(1), 31355. https://doi.org/10.1038/s41598-024-82789-7

Deng, Y., Yuan, W., Ren, E., Wu, Z., Zhang, G., & Xie, Q. (2021). A four-methylated LncRNA signature predicts survival of osteosarcoma patients based on machine learning. Genomics, 113(1), 785-794. https://doi.org/10.1016/j.ygeno.2020.10.010

Wen, B., Chen, J., Ding, T., Mao, Z., Jin, R., Wang, Y., ... & Chen, X. (2024). Development and experimental validation of hypoxia-related gene signatures for osteosarcoma diagnosis and prognosis based on WGCNA and machine learning. Scientific Reports, 14(1), 18734. https://doi.org/10.1038/s41598-024-69638-3

Lin, Z., Zhao, F., Huang, J., Wang, T., & Sun, Z. (2023). The potential for different computed tomography-based machine learning networks to

automatically segment and differentiate pelvic and sacral osteosarcoma from Ewing's sarcoma. Quantitative Imaging in Medicine and Surgery, 13(5), 2892–2904. https://doi.org/10.21037/qims-22-888

Guo, Y., Hu, Y., Yan, Y., Ge, Y., Ma, L., Chen, L., Yao, Y., & Zhou, Z. (2024). MRI-based radiomics for outcome stratification in pediatric osteosarcoma. Frontiers in Oncology, 14, 1387901. https://doi.org/10.3389/fonc.2024.1387901

Caolo, K., Bechar, J., Ferretti, P., & Koskinen, S. K. (2025). An explainable radiomics-based classification model for sarcoma diagnosis. Diagnostics, 15(11), 1432. https://doi.org/10.3390/diagnostics15111432

Kawaguchi, K., Miyama, K., Endo, M., Bise, R., Kohashi, K., Hirose, T., ... & Nakashima, Y. (2024). Viable tumor cell density after neoadjuvant chemotherapy assessed using deep learning model reflects the prognosis of osteosarcoma. npj Precision Oncology, 8(1), 16. https://doi.org/10.1038/s41698-024-00515-y

Ma, H., Zhang, H., Bajgai, J., Rahman, M. H., Pham, T. T., Mo, C., ... & Lee, K. J. (2025). Integrative Single-Cell and Machine Learning Analysis Develops a Glutamine Metabolism–Based Prognostic Model and Identifies MSMO1 as a Therapeutic Target in Osteosarcoma. Biomolecules, 15(12), 1664. https://doi.org/10.3390/biom15121664

Qiao, Y., Yuan, F., Vilalta, A. C., von Eisenhart-Rothe, R., & Hinterwimmer, F. (2025). Assessing the mechanism of osteosarcoma induced by long-term PET exposure: prediction from combined network toxicology, machine learning and molecular docking. Journal of Bone Oncology, 100714. https://doi.org/10.1016/j.jbo.2025.100714

P. J. Adit, Dr.C. Priya, "Cross-Modal Spatial-Semantic Fusion: Integrating MRI and Histopathology for Automated Bone-Sarcoma Severity Assessment," presented at *YUKTI'26: 1st National Conference on INFYNEX*, Chennai, India, May 14-15, 2026.

Ngan, E., Mullikin, D., Theruvath, A. J., Annapragada, A. V., Ghaghada, K. B., Heczey, A. A., & Starosolski, Z. A. (2025). MRI-Based Radiomics for Outcome Stratification in Pediatric Osteosarcoma. Cancers, 17(15), 2586. https://doi.org/10.3390/cancers17152586

Cui, L., Zhao, S., Teng, H. L., Yang, B., Liu, Q., & Qin, A. (2025). Integrins identified as potential prognostic markers in osteosarcoma through multi-omics and multi-dataset analysis. NPJ Precision Oncology, 9(1), 19. https://doi.org/10.1038/s41698-024-00794-5

Wang, Y., Zhang, R., & Liu, H. (2025). End-to-end deep learning prediction of neoadjuvant chemotherapy response in osteosarcoma patients using routine MRI. Journal of Imaging Informatics in Medicine, 38, 1214–1226. https://doi.org/10.1007/s10278-025-01424-7

Liu, F., Wang, Z., Li, B., Wang, D., Liu, M., Gou, F., & Wu, J. (2025). Enhanced digital pathology image recognition via multi-attention mechanisms: the MACC-Net approach. Scientific Reports, 15(1), 31269. https://doi.org/10.1038/s41598-025-17369-4

Bai, Z., Osman, M., Brendel, M., Tangen, C. M., Flaig, T. W., Thompson, I. M., ... & Wang, F. (2025). Predicting response to neoadjuvant chemotherapy in muscle-invasive bladder cancer via interpretable multimodal deep learning. npj Digital Medicine, 8(1), 174. https://doi.org/10.1038/s41746-025-01560-y

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

P. J. Adit, Dr.C. Priya, "BSTCL-OSNet: Bio-Saliency Guided Temporal Capsule Learning for Interpretable Multi-Scale bone-sarcoma Histopathology Classification," in *4th International Conference on Engineering, Social Sciences & Management (ICESSM-2026)*, May 15th-16th, 2026.

Warin, K., Limprasert, W., Paipongna, T., Chaowchuen, S., & Vicharueang, S. (2025). Determination of the oral carcinoma and sarcoma in contrast enhanced CT images using deep convolutional neural networks. Scientific Reports, 15(1), 21672. https://doi.org/10.1038/s41598-025-06318-w

Yadav, D. P., Sharma, B., Noonia, A., & Mehbodniya, A. (2025). Explainable label guided lightweight network with axial transformer encoder for early detection of oral cancer. Scientific Reports, 15(1), 6391.https://doi.org/10.1038/s41598-025-87627-y