

Doi:10.32604/cmc.2025.060163

ARTICLE





Novel Feature Extractor Framework in Conjunction with Supervised Three Class-XGBoost Algorithm for Osteosarcoma Detection from Whole Slide Medical Histopathology Images

Tanzila Saba¹, Muhammad Mujahid¹, Shaha Al-Otaibi², Noor Ayesha³ and Amjad Rehman Khan^{1,*}

¹Artificial Intelligence & Data Analytics Lab, College of Computer & Information Sciences, Prince Sultan University, Riyadh, 11586, Saudi Arabia

²Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, 11671, Saudi Arabia

³Center of Excellence in Cyber Security (CYBEX), Prince Sultan University, Riyadh, 11586, Saudi Arabia

*Corresponding Author: Amjad Rehman Khan. Email: arkhan@psu.edu.sa

Received: 25 October 2024; Accepted: 25 December 2024; Published: 17 February 2025

ABSTRACT: Osteosarcomas are malignant neoplasms derived from undifferentiated osteogenic mesenchymal cells. It causes severe and permanent damage to human tissue and has a high mortality rate. The condition has the capacity to occur in any bone; however, it often impacts long bones like the arms and legs. Prompt identification and prompt intervention are essential for augmenting patient longevity. However, the intricate composition and erratic placement of osteosarcoma provide difficulties for clinicians in accurately determining the scope of the afflicted area. There is a pressing requirement for developing an algorithm that can automatically detect bone tumors with tremendous accuracy. Therefore, in this study, we proposed a novel feature extractor framework associated with a supervised three-class XGBoost algorithm for the detection of osteosarcoma in whole slide histopathology images. This method allows for quicker and more effective data analysis. The first step involves preprocessing the imbalanced histopathology dataset, followed by augmentation and balancing utilizing two techniques: SMOTE and ADASYN. Next, a unique feature extraction framework is used to extract features, which are then inputted into the supervised three-class XGBoost algorithm for classification into three categories: non-tumor, viable tumor, and non-viable tumor. The experimental findings indicate that the proposed model exhibits superior efficiency, accuracy, and a more lightweight design in comparison to other current models for osteosarcoma detection.

KEYWORDS: Medical image processing; deep learning; healthcare; image classification; histopathology

1 Introduction

Osteosarcoma, a high-grade intramedullary sarcoma, unfortunately has a survival rate of less than 30%, making it the most prevalent malignant bone tumor. The prognosis for 70% of patients experiencing local progression has been enhanced with the advent of contemporary chemotherapy treatments. However, more advancement is still needed for individuals with metastasis or recurrence [1]. Society must strive for improved outcomes for those with unfavorable prognoses, given the increasing prevalence of osteosarcoma among young adults. Osteosarcoma is a cancerous tumor that develops from cells responsible for bone development, primarily affecting people in their teenage and early adult years. The condition can arise in any bone structure, primarily in the long shaft of the knee bone and, infrequently, in the upper arm bone [2]. The disease's prognosis has improved as a result of breakthroughs in therapy; nevertheless, patients may encounter



Copyright © 2025 The Authors. Published by Tech Science Press.

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

long-term repercussions owing to powerful medicines. Symptoms of osteosarcoma often arise within a bone, predominantly affecting the femur and tibia bones in the legs and, infrequently, the humeral bone in the arms. Genetic mutations in bone cells cause genetic instructions for fast and widespread cell growth, leading to the development of tumors. Cancer cells can maintain their survival even in the presence of apoptosis, which ultimately results in the growth and destruction of tumors. Cancer cells can separate and spread to various parts of the body, resulting in the development of metastatic cancer [3,4].

Osteosarcoma has substantial variation, featuring variations within individual observations and across various observers, contributing to its diverse nature. However, the rounder shape and more dense packing of osteosarcoma precursor cells contrast with the often more homogeneous size, shape, and density of tumor cells [5]. To accurately evaluate necrosis, it is necessary to consider many histological regions, including areas with hemorrhagic tumors, blood cells, growth plates, nuclear clusters, fibrous tissues, cartilage, osteoclasts, osteoid, osteoblasts, and precursor cells. Recent research [6] has shown the efficacy of Convolutional Neural Network (CNN) in extracting and analyzing data from medical pictures. Using DL and ML methods, research on cancer categorization has shown significant efficacy. Osteosarcoma's intricate pathological characteristics necessitate the expertise of proficient and competent pathologists. On an ongoing basis, each proficient pathologist manages numerous segments [7]. Pathologists must establish an effective decision-making strategy for healthcare informatics to facilitate osteosarcoma analysis and resolve challenges that arise in healthcare facilities. Medical field analysis increasingly employs neural networks due to their superior feature extraction capabilities. This is especially true for tasks like MRI segmentation of osteosarcoma and supplemental staging of lung cancer [8]. CNN is primarily used for automated data extraction from image data [9].

Osteosarcoma slides serve as a crucial diagnostic tool for cancer and are commonly utilized with noninvasive imaging techniques like MRI and CT [10]. These approaches facilitate the identification of cancerous areas and provide quantitative assessments for monitoring treatment responses and planning surgical procedures. Nevertheless, using tissue slides to examine and diagnose physically osteosarcoma patients might be subjective and time-consuming. Utilizing whole slide image (WSI) analysis can enhance the reliability of the study [11]. This methodology may obtain further information from tissue slides, resulting in a more correct diagnosis. It is planned for an automated method to use morphological and contextual signals from digital whole slide images (WSIs) to find osteosarcoma slides based on a histological examination [12]. This approach facilitates the utilization of image processing and analysis methodologies, enhancing the precision and reliability of cancer diagnosis. Osteosarcoma is one malignancy that is extremely variable and susceptible to numerous interpretations. Specific precursor cells and tumor cells in osteosarcoma exhibit comparable blue staining patterns; however, tumor cells are more asymmetrical, densely clustered, and spherical than precursor cells [13,14]. Many histological regions must be considered when determining the percentage of necrosis with precision. Recent medical research provides empirical evidence that deep learning can efficiently extract and analyze information from medical images. The following are main contributions of this work:

- We proposed a novel feature extractor framework with a supervised three-class XGBoost algorithm for detecting osteosarcoma in whole-slide histopathology images. This method allows for quicker and more effective data analysis.
- To address the imbalanced dataset problem, we used oversampling techniques such as ADASYN and SMOTE.
- To enhance the dataset size, we used several augmentation techniques. After oversampling and augmentation techniques, non-tumor cases increased from 536 to 10,538, non-viable tumor cases from 263 to 10,497, and viable tumor cases from 345 to 10,512.

• During the training phase, the model is exposed to diverse augmented images, enhancing its resistance to various data orientations, lighting conditions, and transformations, thereby reducing overfitting and improving performance.

2 Literature Review

Histology images of osteosarcoma were put through transfer learning techniques, which use convolutional neural networks that have already been trained to tell the difference between pictures of dead tissue and pictures of healthy tissue. A multitude of methodologies were utilized to preprocess and classify the dataset. Whole slide images were employed during training to improve the precision of transfer learning models such as VGG-19 and Inception V3. The models address difficulties associated with classifying multiple classes and binary data. Accuracy-wise, the VGG-19 model surpasses all other models. Histologic images show that the enhanced model is remarkably effective at precisely identifying the malignant nature of osteosarcoma [15].

The study [16] wanted to show how different types of osteosarcoma are and how they react to chemotherapy by using 13 machine-learning models and 40 scanned slide images to sort the tissue into three groups: viable, necrotic, and non-tumorous. The SVM model was chosen based on its exceptional performance assessment score. In addition, a deep learning architecture was created and trained using the same dataset. The ROC was computed to distinguish between non-tumorous and tumorous areas. In addition, conditional discrimination was performed to differentiate between live and necrotic tumors. The models demonstrated exceptional performance. Using the trained models, researchers detected regions of interest on image tiles obtained from whole slide test images. A tumor-prediction map was created to demonstrate the proportional presence of viable and necrotic tumors compared to the slide image. Due to its challenging diagnosis, osteosarcoma is an uncommon but frequently benign bone cancer that family practitioners see frequently. Classifying histological images into viable, non-viable, and non-tumor categories is a common challenge for pathologists. Osteosarcoma was effectively and reliably classified into these three groups using the Random Forest machine learning method, which yielded an AUC value of 0.95 and an accuracy of 92.40% [17].

Utilizing CNN can greatly improve the effectiveness and precision of classifying osteosarcoma tumors into several categories, such as VTs, necrosis, and non-tumorous regions [18]. Kawaguchi et al. [19] developed density cell chemotherapy through deep methods for cancer detection. The ResNet101 supervised deep learning model was selected for bone cancer identification, with a prediction accuracy of 90.36% and a precision of 89.5%. The model's efficacy in diagnosing bone cancer was successfully proven using user input weighting, showcasing its capacity to correctly provide anticipated results [20].

The DL methodology was employed to classify osteosarcoma cells into discrete groups, a prominent manifestation of bone cancer that mostly affects children and young persons. Mesenchymal Stromal Cells (MSCs) were employed to cultivate several cell types, which undergo the process of differentiation into osteoblasts and osteosarcoma cells. The specimens were imaged using an optical microscope, and deep learning algorithms were employed to identify and organize individual cells, resulting in a rating accuracy [21]. The paper presents an innovative ensemble learning model to diagnose bone malignancies. Their model incorporates a pre-processing technique that eliminates undesirable regions, reduces noise, and enhances image quality. The SIKC method specifies the segmentation process, while the RGB histogram and spatial GLDM extract texture and color features. The model additionally conducts a severity analysis and categorizes images according to viable, non-viable, or non-tumorous malignancies [22]. It was possible to make a better system for classifying osteosarcoma tumors by combining CNN-based architectures with the multilayer perceptron (MLP) method. The platform uses CNN models that have already been trained to pick out important features and get valuable data from whole slide imaging images. Removing less-important features improves the overall ability of the model to make correct predictions. The model outperformed previous

approaches, achieving 95.2% accuracy in multiclass classification. This model can aid medical practitioners in diagnosing osteosarcoma and provide current prognostications by connecting with an online service and employing the FastAPI web framework. Theyy tested how stable the model was using a modid MLP classifier for binary andiclass osteosarcoma classification within the framework of five-fold cross-validation [23].

Using machine learning techniques in medical instruments was a noteworthy development in medical technology. To predict osteosarcoma survival, three crucial machine learning techniques are compared in work [24]. The higher efficacy of SVM and ANN when red to random forests demonstrates their potential to provide more precise survival predictions. A comparative study was performed on regularly employed methodologies employing feature and image datasets to create a prediction system for clinical settings. The feature dataset employed techniques such as Extra Trees Classifier, Convolutional Neural Network, XGBoost Classifier, and DenseNet classifier. The fusion of viable and necrotic tumors resulted in a maximum training accuracy of 96.22%, while the non-tumor vs. viable group attained a slightly lower accuracy of 94.56%. The convolutional neural networks were utilized to examine the image collection for osteosarcoma forecasts, showcasing their diminished intricacy. The model was effectively developed with 2.5 million trainable parameters [25]. Vezakis et al. [10] presented a methodological technique that involves comparison. They assessed several pre-trained models using transfer learning to standardize and resize input photographs based on the dimensions of each model. The MobileNetV2 model achieved 91.0% accuracy, making it the most precise.

3 Materials and Methods

The work flow of the proposed framework, dataset description, organisation, preprocessing, enhancement via augmentation and oversampling approaches, and the deep and proposed framework are evaluated using cross-validation, ANOVA test and other comparative analyses is presented in Fig. 1.



Figure 1: The work flow of the proposed framework, dataset description, organization, preprocessing, enhancement via augmentation and oversampling approaches, and the deep and proposed framework are evaluated using cross-validation, ANOVA test and other comparative analyses

3.1 Preprocessing and Data Augmentation

In machine learning, scaling images ensures that models receive uniform data, enabling them to generalize effectively across various dimensions. The consistency of the data allows models to focus on consistent patterns, extracting key features and patterns even when there are differences in scale. Following preprocessing steps are adopted in this work:

- The deep learning models require fixed-size input data, so we convert all the images into fixed input shapes (224×224). Ensuring a uniform image size also streamlines computations and minimizes memory demands, improving training effectiveness.
- Normalization ensures that the brightness and contrast levels between pixels in an image are more uniform. Most 8-bit images range from 0 to 255, where 255 are white and 0 are black. Normalization is a useful tool for improving images' sharpness or ensuring that all pixels have the same value before processing them further. Normalizing or standardizing color channels can lessen the impact of different colors [26].

Data augmentation is an essential technique for machine learning, especially when working with image data. The procedure entails altering the existing data, enhancing variability and the model's capacity to generalize across diverse contexts. This reduces the problem of overfitting, which occurs when a model is so skilled at recalling the training data that it is difficult to adjust for new examples. During training, the model is equipped with new data that helps it recognize patterns in various sizes, shapes, locations, textures and lighting conditions. Improving the model's adaptability makes it more sensitive and sensitive to small changes, thus being better prepared to deal with real-world situations. Data augmentation techniques including flipping, distortion, compression, resampling, distortion, and noise reduction are used to simulate real-world changes in image properties. This improves the robustness of the model and ensures its safety and performance in practical situations [15]. Dataset classes using oversampling and augmentation techniques are shown in Table 1.

Class label	Primary	Final 1	Final 2	Final 3
NT	536	10508	10508	10538
VT	345	5863	10508	10512
NVT	263	5415	10508	10497
Total	1144	21786	21786	31524

Table 1: Dataset classes using oversampling and augmentation techniques

3.2 SMOTE vs. ADASYN

ADASYN and SMOTE are re-sampling techniques employed to rectify class imbalances in deep learning tasks related to object detection and image classification. SMOTE minimizes the losses of models of majority and minority classes by generating new instances that alternate between scenarios. In contrast, ADASYN exhibits its versatility by generalizing the reconstruction process to correspond with the number of nodes in the local vicinity, thereby enhancing the efficiency of addressing significant disparities by focusing on regions with smaller clusters. ADASYN provides a more practical approach for managing sequence imbalances by dynamically adjusting its alignment strategy, which improves the capacity to replicate and optimize various image segmentation challenges [27,28]. Table 2 shows the parameters setting for SMOTE and ADASYN.

SMOTE	ADASYN	Library
Sampling_strategy = "minority"	Sampling_strategy = "minority"	
K_neighbors = 5	$N_neighbors = 5$	Imbalanced learn
Random_state = 42	Random_state = 42	

Table 2: Parameter setting for SMOTE and ADASYN

3.3 Proposed Method

VGG16 was created by the Visual Genomics Group (VGG) at the University of Oxford for the ILSVRC (ImageNet Large Scale Visual Recognition Challenge). The design included sixteen layers, thirteen of which were duplicated and three were completely collaborative. The primary aim of this research is to assess the efficacy of the CNN function. The VGG team used a straightforward modelling strategy with a 3×3 filter configuration and a gradient filter design. The objective was to enhance efficiency by augmenting the model's complexity. The first two convolutional layers employ a stride of 1 and a small receptive field of 3×3 . Thanks to these layers, the network can analyze and represent simple visual features in the input image, such as edges, textures, and basic patterns. Max-pooling layers are applied after each set of convolutional layers, with a stride of 2 and a window size of 2×2 . This design decision is intended to reduce processing costs, achieve translation invariance, and lessen the spatial dimensions of the input.

The convolutional layers that follow, ranging from 5 to 10, gradually improve in depth and intricacy to capture increasingly complex characteristics. Augmenting the quantity of convolutional layers amplifies the model's ability to obtain hierarchical representations of the input image. This step applies max pooling after every two convolutional layers to decrease spatial dimensionality and improve abstraction. Three layers comprise the VGG16 architecture, and a convolutional layer connects them. This convolutional layer uses the recovered features to classify by combining spatial information from the layers that came before it. The number of classes in the ImageNet dataset during the model's development is reflected in the 1000 neurons that make up the output layer of the VGG16 model. The layer uses the softmax activation function to translate the network's raw output into probabilities for each class. Fig. 2 showcases a feature extractor framework integrated with a supervised three-class XGBoost algorithm for detecting Osteosarcoma.

The VGG16 architecture is widely acknowledged for its straightforwardness and reliability. The network is easily understandable and trainable because it frequently utilizes small 3×3 filters and multiple layers of convolution. The simple design has served as a source of inspiration for creating more buildings. VGG16 STATEs that there is a correlation between the depth of a network and its capacity to acquire complex characteristics. This link ultimately results in enhanced performance in the classification of images. This finding has significant implications for the progress of deep neural networks. To make the model specific to the given job, we can use transfer learning with VGG16 and fine-tune the model's parameters using a smaller dataset.

To meet the demands of the current task, the fully connected layers are adjusted, while the convolutional layers, which are responsible for extracting features, retain their understanding of prominent visual characteristics. This strategy demonstrates particular benefits when labeled data are scarce for the new task. This result can be attributed to the model's ability to utilize the knowledge gained from ImageNet pre-training. One major advantage of using transfer learning with VGG16 is the huge decrease in training time and data needs. After training, the model can enhance its performance on a particular job using fewer labeled data points. This is possible since the model has gained various hierarchical characteristics from a diverse dataset

like ImageNet. Therefore, it offers an efficient application when obtaining extensive labeled datasets that are excessively costly or challenging.



Figure 2: Novel feature extractor framework in conjunction with supervised three class-XGBoost algorithms for osteosarcoma detection

We selected GG16 as the feature extractor due to its popularity as a robust convolutional neural network architecture, recognised for its efficacy in extracting spatial data and intricate structures from images. The meticulously crafted VGG16 model on the ImageNet dataset has shown robust generalisation across several domains, including medical imaging. The convolutional layers may extract significant features, including corners, edges, and high-dimensional semantics, essential for differentiating characteristics in structured datasets. We augment the series' complexity by examining its foundation and using its characteristics as feature vectors, utilising deep learning with pre-trained models.

XGBoost was classified for its precision, scalability, and versatility. This is especially beneficial for managing graphical data derived from deep learning tasks. The XGBoost gradient boosting approach facilitates the model's ability to learn intricate nonlinear decision boundaries, making it appropriate for medical image applications. XGBoost produces feature significance curves that enhance the interpretability of classification outcomes, which is crucial for clinical applications.

The integration of VGG16 with XGBoost leverages the advantages of both models VGG16 is proficient in feature extraction from low-dimensional image data, while XGBoost is tailored for classification in structured environments. This method enables us to reconcile efficiency with computational complexity, assuring the model's effective generalisation to novel data.

4 Results and Discussion

This section presents the experiments and evaluation procedures used in the study to analyze the effectiveness of the proposed framework.

4.1 Experimental Setup

This study focuses on detecting osteosarcoma bone tumors using deep learning and a unique feature extractor based on the XGBoost algorithm. The experiments were conducted using Python programming and following specific parameters. A data splitting method called hold-out was used, with 90% of images designated for training and 10% for testing. Three different pre-processing methods were used to prepare the input images. For training the deep transfer learning models, each image was resized to 224 pixels by 224 pixels, with a batch size of 32. Optimization was performed using the Adam algorithm, with a learning rate of 0.0001. The models were trained independently, and the results from the recommended model were integrated through cross validation. The deep learning backend consisted of Keras and TensorFlow, and the NVIDIA GeForce RTX 2060 graphics card was used for training and testing. Accuracy is determined using true positive tumor predictions divided by total number of predictions. Precision determined using true positive tumor prediction divided by true positive tumor plus false positive tumor predictions. Recall determined using true positive tumor prediction divided by true positive tumor plus false negative tumor predictions. F1 score is the mean of precision and recall. The F1 score is a critical deep learning metric that accurately assesses the precision and recall of a model.

$$Accuracy = \frac{TPT + TNT}{TPT + TNT + FPT + FNT}$$
(1)

$$Precision = \frac{TPT}{TPT + FPT}$$
(2)

$$\operatorname{Recall} = \frac{\operatorname{TPT}}{\operatorname{TPT} + \operatorname{FNT}}$$
(3)

F1 score =
$$2 \times \frac{\text{Precsion} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
 (4)

4.2 Dataset Information

The study evaluated a dataset of histological images of osteosarcoma stained with hematoxylin and eosin. The data was obtained from archived samples from fifty pediatric patients who had treatment at the Children's Medical Center in Dallas from 1995 to 2015. The images were categorized as non-tumor, viable tumor, or non-viable tumor, depending on the predominant cancer types. The collection comprises 1144 images, each with a dimension of 1024 by 1024 pixels. The allocation of these images is as follows: Out of the total, 536 cases (47%) do not have tumors, 263 cases (23%) have necrotic components, and 345 cases (30%) have active tumors. The dataset is publicly accessible and categorized into three groups: non tumor, viable tumor, and non-viable tumor [29].

Nevertheless, the quantity of images available was inadequate to meet the demands for training and testing. Therefore, data augmentation and sampling methods were employed to enhance the data. Regarding classification, the most important criteria for evaluation are accuracy, precision, recall, and F1 score.

4.3 Results Using Primary Dataset

The experimental results using the primary dataset mean on the original dataset are presented in Table 3. CNN model achieved 83.18% accuracy overall, 81.48% precision for non-tumor class, 86.08% for non-viable tumor class, and 81.48% F1 score for non-tumor, while VGG-19 achieved more positive results than CNN and MobileNet, with 87.24% overall accuracy on the primary dataset. The proposed model achieved 90.14% overall accuracy, a 90.38% F1 score, and 94.01% precision for the non-tumor class. All models achieved a lower score for the viable tumor class.

Table 3: Performance of transfer learning and proposed feature based algorithm for osteosarcoma detection using primary dataset

Models	Accuracy	Label	Precision	Recall	F1
Primary-CNN	83.18	NT	81.48	81.48	81.48
		VT	67.56	71.42	69.44
		NVT	86.08	70.83	65.38
Primary-ResNet-50	86.08	NT	88.23	83.33	85.71
		VT	74.28	74.28	74.28
		NVT	68.96	76.92	72.72
Primary-MobileNet	80.86	NT	73.58	72.22	72.89
		VT	83.33	71.42	76.92
		NVT	56.25	69.26	62.06
Primary-VGG-19	87.24	NT	83.33	83.33	83.33
		VT	81.81	77.14	79.41
		NVT	75.01	80.76	77.77
Primary-Proposed	90.14	NT	94.01	87.03	90.38
		VT	85.29	82.85	84.08
		NVT	70.96	84.61	77.19

4.4 Results Using Final 1 Dataset

The results obtained from conducting experiments using the Final 1 dataset are shown in Table 4. The CNN model obtained an overall accuracy of 87.05%, a precision of 83.02% for the non-tumor class, an accuracy of 87.10% for the non-viable tumor class, and an F1 score of 83.84% for the non-tumor class. On the other hand, the VGG-19 model outperformed both the CNN and MobileNet models, with an overall accuracy of 88.92% on the main dataset. The suggested model attained an overall accuracy of 94.18% and an F1 score of 95.08%. Additionally, it demonstrated a precision of 96.38% for the non-tumor class. MobileNet had the lowest results, whereas VGG-19 received the greatest results.

Table 4: Performance of transfer learning and proposed feature based algorithm for osteosarcoma detection using Final 1 dataset

Models	Accuracy	Label	Precision	Recall	F1
Final 1-CNN	87.05	NT	83.02	84.68	83.84
		VT	71.60	77.47	74.42

(Continued)

ble 4 (continued)					
Models	Accuracy	Label	Precision	Recall	F1
		NVT	87.10	76.01	81.18
Final 1-ResNet-50	87.30	NT	83.52	84.87	84.19
		VT	71.96	77.98	74.85
		NVT	87.18	76.56	81.53
Final 1-MobileNet	85.86	NT	83.20	82.96	83.08
		VT	66.91	76.27	71.06
		NVT	86.71	73.43	79.52
Final 1-VGG-19	88.92	NT	89.59	82.77	86.08
		VT	76.42	80.20	78.26
		NVT	80.43	88.01	84.05
Final 1-Proposed	94.18	NT	96.38	93.81	95.08
		VT	86.30	86.00	86.15
		NVT	87.23	92.06	89.58

4.5 Results Using Final 2 Dataset

Table 5 displays the results of running experiments using the Final 2 dataset. The VGG-19 model achieved an overall accuracy of 92.91%, a precision of 91.62% for the non-tumor class, an accuracy of 89.78% for the non-viable tumor class, and an F1 score of 90.56% for the non-tumor class. The CNN and MobileNet models were not as successful as the VGG-19 model, which performed much better. With an accuracy rate of 91.60% across the board, the MobileNet model has the lowest. The proposed model achieved a total accuracy of 97.25% and received an F1 score of 95.61%. Furthermore, it displayed an accuracy of 94.85% for the class of tumors that were not present. The results that MobileNet obtained were the worst, while the results that VGG-19 received were the lowest.

Table 5: Performance of transfer learning and proposed feature based algorithm for Osteosarcoma detection using Final2 dataset

Models	Accuracy	Label	Precision	Recall	F1
Final 2-CNN	92.76	NT	88.39	89.15	88.77
		VT	88.80	89.81	89.30
		NVT	90.92	88.48	89.30
Final 2-ResNet-50	92.38	NT	87.81	88.48	88.15
		VT	87.89	89.15	88.52
		NVT	90.07	88.10	89.08
Final 2-MobileNet	91.60	NT	85.56	86.86	86.21
		VT	89.31	87.44	88.36
		NVT	87.41	87.91	87.66
Final 2-VGG-19	92.91	NT	91.62	89.53	90.56
		VT	86.85	89.91	88.35
		NVT	89.78	88.67	89.22

(Continued)

Models	Accuracy	Label	Precision	Recall	F1
Final 2-Proposed	97.25	NT	94.85	96.38	95.61
		VT	96.62	95.33	95.97
		NVT	96.18	95.90	96.04

Table 5 (continued)

4.6 Results Using Final 3 Dataset

Table 6 presents the experimental findings using the Final 3 dataset. Overall, the ResNet-50 model got 92.35 percent correct in the Final 3 dataset. It got 87.81% correct for the non-tumor class, 89.17% for the non-viable tumor class, and 88.89% for the non-tumor F1 score. On the other hand, VGG-19 performed better than ResNet-50 and MobileNet, with 91.17% overall accuracy. The suggested model obtained 93.86% precision for the non-tumor class, 94.78% F1 score, and overall accuracy of 96.84%.

Table 6: Performance of transfer learning and proposed feature based algorithm for osteosarcoma detection using Final3 dataset

Models	Accuracy	Label	Precision	Recall	F1
Final 3-CNN	92.70	NT	88.00	89.27	88.63
		VT	89.00	89.42	89.21
		NVT	90.20	88.49	89.33
Final 3-ResNet-50	92.35	NT	87.81	88.23	88.02
		VT	88.07	89.24	88.65
		NVT	89.17	88.09	88.89
Final 3-MobileNet	91.48	NT	85.29	86.43	85.86
		VT	89.12	87.34	88.22
		NVT	87.32	87.90	87.61
Final 3-VGG-19	92.96	NT	91.17	89.18	90.16
		VT	86.96	90.10	88.50
		NVT	90.33	89.04	89.68
Final 3-Proposed	96.84	NT	93.86	95.73	94.78
		VT	95.58	94.86	95.22
		NVT	96.43	95.23	95.83

4.7 Training and Validation Curves

Training and validation accuracy are crucial in deep learning models, particularly during training. Training accuracy measures the model's performance on the training dataset, demonstrating its capacity to acquire knowledge and correctly represent the data. Validation accuracy measures the model's effectiveness using a separate validation dataset, evaluating its ability to apply knowledge to new, unfamiliar data. The difference between training and validation accuracy lies in their goals and implications. Fig. 3 illustrates the accuracy curves for training and validation using four datasets. Fig. 3a illustrates the accuracy of the training and validation sets using the primary (original) dataset. The best accuracy was attained at epoch 26 for the training set and at epoch 22 for the validation set. Fig. 3b displays the accuracy of the training and validation



sets using the Final 1 dataset. The maximum accuracy was attained at epoch 30 for the training set and epoch 27 for the validation set. Fig. 3c displays the accuracy of the training and validation sets using the Final 2 dataset. Fig. 3d displays the accuracy of the training and validation sets using the Final 3 dataset.

Figure 3: The proposed approach demonstrates the train and validation accuracy and the best epoch for all datasets and highlights the epochs

4.8 AUC ROC Curves

ROC AUC is a reliable evaluation metric even in imbalanced datasets, making it useful for choosing models, adjusting hyperparameters, and comparing training approaches or architectures. Deep learning models are used for reliable and effective performance in various applications. CNN attained an AUC of 0.9212, ResNet-50 at 0.9318, MobileNet at 0.9162, and VGG-19 at 0.9282. Using the primary dataset, the proposed model attained an AUC score of 0.9426; the curves are illustrated in Fig. 4a.

CNN had an AUC of 0.9321, ResNet-50 got 0.9412, MobileNet got 0.9273, VGG-19 got 0.9542, and the proposed model got 0.9676 when used on the Final 1 dataset as shown in Fig. 4b.



Figure 4: The proposed approach showcases the ROC-AUC for all datasets

As shown in Fig. 4c, CNN attained an AUC of 0.9512, ResNet-50 attained 0.9461, MobileNet attained 0.9302, VGG-19 attained 0.9523, and the proposed model attained a 0.9910 AUC when applied to the Final 2 dataset. VGG-19 attained an AUC of 0.9551, while CNN attained 0.9501, ResNet-50 attained 0.9432, and MobileNet attained 0.9403. Using the Final 3 dataset, the proposed model attained an AUC score of 0.9863; the curves are illustrated in Fig. 4d.

4.9 Comparison with State of the Art Techniques

Comparing the proposed approach with the state-of-the-art techniques is important to check the efficacy for osteosarcoma detection. The study examines the efficacy of deep learning models. It demonstrates that the proposed method surpasses conventional approaches such as support vector machines (94.56%) and random forests (92.40%) in terms of performance as shown in Table 7. The accuracy rate of the model is 96.84%, which is higher than the accuracy rates of the deep learning architectures MobileNetV2 (91.01%) and ResNet-101 (90.36%). This suggests that the MLP approach is proficient at identifying patterns in data and can serve as a dependable remedy to existing issues, showcasing its reliability. Gyasi-Agyei et al. [26] used logistic

regression and principal component analysis to diagnose osteosarcoma, but their results were very poor (77.52% accuracy). Nabid et al. [30] developed a deep learning recurrent CNN technique to detect the cancer in bones and attained some better 89% accuracy. Ahmed et al. [31] also utilized CNN technique for same dataset and attained 86.17% accuracy. The previous techniques did not perform well because some studies did not properly utilize oversampling, augmentation and extraction techniques. The proposed approach achieved outclasses performance using a balanced dataset.

Ref.	Employed dataset	Technique	Accuracy
Vezakis et al. [10]	Medical histopathology	MobileNetV2	91.01
Arunachalam et al. [16]	Medical histopathology	DLA	93.30
Mahore et al. [17]	Medical histopathology	RF	92.40
Gawade et al. [20]	Medical histopathology	ResNet-101	90.36
Aziz et al. [23]	Medical histopathology	MLP	95.23
Srivastava et al. [25]	Medical histopathology	SVM	94.56
Gyasi-Agyei et al. [26]	Medical histopathology	LR+PCA	77.52
Nabid et al. [30]	Medical histopathology	RCNN	89.00
Ahmed et al. [31]	Medical histopathology	CNN	86.17
[Proposed]	Medical histopathology	VGG15+XGBoost	96.84

Table 7: Comparison with state of the art techniques

5 Conclusion

Using computer-aided imaging techniques for automated histological image classification is paramount in medical image processing. Examining histology photographs under a microscope takes a lot of time and a significant amount of commitment. Within the field of histology, automated diagnosis enables pathologists to devote more time and effort to cases that are of critical importance. The goal of this study was to enable the diagnosis of osteosarcoma in whole-slide histopathology images by integrating a supervised threeclass XGBoost algorithm with our own custom-built feature extractor framework. SMOTE and ADASYN oversampling techniques augment and balance the dataset. The overfitting issue is solved after balancing the imbalanced samples of the bone tumor dataset. The model achieved 96.84% accuracy and, 95.22% F1 score for viable bone tumors. The XGBoost classifier uses the extracted features from the VGG16 architecture to classify osteosarcoma cancer. Experiments demonstrate that a three-class XGBoost classifier works best using deep-extracted features. The results of the experiments show that the proposed model for osteosarcoma detection is superior to the currently considered to be state-of-the-art in terms of accuracy, efficiency, and lightweight design. Despite the strengths, the study has some limitations:

- Oversampling methods are only relevant to training data. In imbalanced data, which often occurs in real-world applications, model performance indicators may be biased towards the majority class.
- Oversampling may exacerbate mistakes in noisy or misclassified outclassed samples by producing more labels, diminishing model performance.

In the future, we intend to integrate several osteosarcoma bone cancer datasets with identical disease classifications and devise a more robust ensemble methodology using the autoencoder technology. A prototype will be developed for the dependable and precise identification of osteosarcoma.

Acknowledgement: This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors also thank AIDA Lab CCIS, Prince Sultan University, Riyadh, Saudi Arabia for support.

Funding Statement: This research is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Conceptualization: Amjad Rehman Khan, Tanzila Saba, Muhammad Mujahid, Shaha Al-Otaibi; methodology, Noor Ayesha, Shaha Al-Otaibi; software: Muhammad Mujahid, Amjad Rehman Khan; validation: Noor Ayesha, Amjad Rehman Khan, Muhammad Mujahid; writing—original draft preparation: Amjad Rehman Khan, Noor Ayesha, Muhammad Mujahid; writing—review and editing: Amjad Rehman Khan; visualization, Shaha Al-Otaibi; supervision: Amjad Rehman Khan, Tanzila Saba, Shaha Al-Otaibi; project administration: Tanzila Saba, Shaha Al-Otaibi. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The dataset is publicly available at: Leavey P, Sengupta A, Rakheja D, Daescu O, Arunachalam HB, and Mishra R. (2019). Osteosarcoma Data from UT Southwestern/UT Dallas for Viable and Necrotic Tumor Assessment (Osteosarcoma-Tumor-Assessment) [Data set]. The Cancer Imaging Archive. DOI: 10.7937/tcia. 2019.bvhjhdas.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- 1. Choi JH, Ro JY. The 2020 WHO classification of tumors of bone: an updated review. Adv Anat Pathol. 2021;28(3):119-38. doi:10.1097/PAP.0000000000293.
- 2. Chou AJ, Geller DS, Gorlick R. Therapy for osteosarcoma: where do we go from here? Pediatr Drugs. 2008;10(5):315–27. doi:10.2165/00148581-200810050-00005.
- 3. Alshahrani H, Sharma G, Anand V, Gupta S, Sulaiman A, Elmagzoub MA, et al. An intelligent attention-based transfer learning model for accurate differentiation of bone marrow stains to diagnose hematological disorder. Life. 2023;13(10):2091. doi:10.3390/life13102091.
- 4. Omar Bappi J, Rony MAT, Shariful Islam M, Alshathri S, El-Shafai W. A novel deep learning approach for accurate cancer type and subtype identification. IEEE Access. 2024;12:94116–34. doi:10.1109/ACCESS.2024.3422313.
- 5. de Andrea CE, Petrilli AS, Jesus-Garcia R, Bleggi-Torres LF, Alves MTS. Large and round tumor nuclei in osteosarcoma: good clinical outcome. Int J Clin Exp Pathol. 2011;4(2):169.
- Spanhol FA, Oliveira LS, Petitjean C, Heutte L. Breast cancer histopathological image classification using convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN); 2016; Vancouver, BC, Canada: IEEE. p. 2560–7. doi:10.1109/IJCNN.2016.7727519.
- Wu J, Yang S, Gou F, Zhou Z, Xie P, Xu N, et al. Intelligent segmentation medical assistance system for MRI images of osteosarcoma in developing countries. Comput Math Methods Med. 2022;2022(6):7703583. doi:10.1155/2022/ 7703583.
- Mohan BC. Osteosarcoma classification using multilevel feature fusion and ensembles. In: 2021 IEEE 18th India Council International Conference (INDICON); 2021; Guwahati, India: IEEE. p. 1–6. doi:10.1109/INDICON52576. 2021.9691543.
- El-Shafai W, Mahmoud AA, Ali AM, El-Rabaie ESM, Taha TE, El-Fishawy AS, et al. Efficient classification of different medical image multimodalities based on simple CNN architecture and augmentation algorithms. J Opt. 2024;53(2):775–87. doi:10.1007/s12596-022-01089-3.
- 10. Vezakis IA, Lambrou GI, Matsopoulos GK. Deep learning approaches to osteosarcoma diagnosis and classification: a comparative methodological approach. Cancers. 2023;15(8):2290.

- 11. Ni YL, Zheng XC, Shi XJ, Xu YF, Li H. Deep convolutional neural network based on CT images of pulmonary nodules in the lungs of adolescent and young adult patients with osteosarcoma. Oncol Lett. 2023;26(2):1–8.
- 12. Baidya Kayal E, Ganguly S, Sasi A, Sharma S, Ds D, Saini M, et al. A proposed methodology for detecting the malignant potential of pulmonary nodules in sarcoma using computed tomographic imaging and artificial intelligence-based models. Front Oncol. 2023;13:1212526.
- 13. Xu R, Tang J, Li C, Wang H, Li L, He Y, et al. Deep learning-based artificial intelligence for assisting diagnosis, assessment and treatment in soft tissue sarcomas. Meta-Radiol. 2024;2(2):100069.
- Pan L, Wang H, Wang L, Ji B, Liu M, Chongcheawchamnan M, et al. Noise-reducing attention cross fusion learning transformer for histological image classification of osteosarcoma. Biomed Signal Process Control. 2022;77(1):103824. doi:10.1016/j.bspc.2022.103824.
- 15. Anisuzzaman DM, Barzekar H, Tong L, Luo J, Yu Z. A deep learning study on osteosarcoma detection from histological images. Biomed Signal Process Control. 2021;69(5):102931. doi:10.1016/j.bspc.2021.102931.
- Arunachalam HB, Mishra R, Daescu O, Cederberg K, Rakheja D, Sengupta A, et al. Viable and necrotic tumor assessment from whole slide images of osteosarcoma using machine-learning and deep-learning models. PLoS One. 2019;14(4):e0210706. doi:10.1371/journal.pone.0210706.
- Mahore S, Bhole K, Rathod S. Machine learning approach to classify and predict different osteosarcoma types. In: 2021 8tth International Conference on Signal Processing and Integrated Networks (SPIN); 2021; Noida, India: IEEE. p. 641–5. doi:10.1109/SPIN52536.2021.9566061.
- Mishra R, Daescu O, Leavey P, Rakheja D, Sengupta A. Histopathological diagnosis for viable and non-VT prediction for osteosarcoma using convolutional neural network. In: Bioinformatics Research and Applications: 13th International Symposium, ISBRA 2017; 2017 May 29–Jun 2; Honolulu, HI, USA: Springer International Publishing. p. 12–23. doi:10.1007/978-3-319-59575-7_2.
- 19. Kawaguchi K, Miyama K, Endo M, Bise R, Kohashi K, Hirose T, et al. VT cell density after neoadjuvant chemotherapy assessed using deep learning model reflects the prognosis of osteosarcoma. npj Precis Oncol. 2022;6(1):1–10.
- Gawade S, Bhansali A, Patil K, Shaikh D. Application of the convolutional neural networks and supervised deeplearning methods for osteosarcoma bone cancer detection. Healthcare Anal. 2023;3(6):100153. doi:10.1016/j.health. 2023.100153.
- D'Acunto M, Martinelli M, Moroni D. Deep learning approach to human osteosarcoma cell detection and classification. In: Multimedia and Network Information Systems: Proceedings of the 11th International Conference MISSI; 2018; Wrocław, Poland: Springer International Publishing. p. 353–61. doi:10.1007/978-3-319-98678-4_36.
- 22. Deepak KV, Bharanidharan R. Osteosarcoma detection in histopathology images using ensemble machine learning techniques. Biomed Signal Process Control. 2023;86(1):105281. doi:10.1016/j.bspc.2023.105281.
- 23. Aziz MT, Mahmud SH, Elahe MF, Jahan H, Rahman MH, Nandi D, et al. A novel hybrid approach for classifying osteosarcoma using deep feature extraction and multilayer perceptron. Diagnostics. 2023;13(12):2106. doi:10.3390/ diagnostics13122106.
- Muthaiyah S, Singh VA, Zaw TOK, Anbananthen KS, Park B, Kim MJ. A binary survivability prediction classification model towards understanding of osteosarcoma prognosis. Emerg Sci J. 2023;7(4):1294–314. doi:10. 28991/ESJ-2023-07-04-018.
- Srivastava DK, Batta A, Gupta T, Shukla A. Prediction of osteosarcoma using machine learning techniques. In: Proceedings of 3rd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2022; 2023; Singapore: Springer Nature. p. 469–80.
- 26. Gyasi-Agyei A, Al-Quraishi T, Das B, Agbinya JI. Exploratory analysis and preprocessing of dataset for the classification of osteosarcoma types. In: Proceedings of International Conference for ICT (ICICT); 2023; Zambia. Vol. 5, p. 36–43.
- Pereira HM, Leite Duarte ME, Ribeiro Damasceno I, de Oliveira Moura Santos LA, Nogueira-Barbosa MH. Machine learning-based CT radiomics features for the prediction of pulmonary metastasis in osteosarcoma. Br J Radiol. 2021;94(1124):20201391. doi:10.1259/bjr.20201391.

- 28. Bai BL, Wu ZY, Weng SJ, Yang Q. Application of interpretable machine learning algorithms to predict distant metastasis in osteosarcoma. Cancer Med. 2023;12(4):5025–34. doi:10.1002/cam4.5225.
- 29. Leavey P, Sengupta A, Rakheja D, Daescu O, Arunachalam HB, Mishra R. Osteosarcoma data from UT Southwestern/UT dallas for viable and necrotic tumor assessment (Osteosarcoma-tumor-assessment) [Data set]. Cancer Imag Arch. 2019. doi:10.7937/tcia.2019.bvhjhdas.
- Nabid RA, Rahman ML, Hossain MF. Classification of osteosarcoma tumor from histological image using sequential RCNN. In: 2020 11th International Conference on Electrical and Computer Engineering (ICECE); 2020; Dhaka, Bangladesh. p. 363–6. doi:10.1109/ICECE51571.2020.9393159.
- 31. Ahmed I, Sardar H, Aljuaid H, Khan FA, Nawaz M, Awais A. Convolutional neural network for histopathological osteosarcoma image classification. Comput Mater Contin. 2021;69(3):3365–81. doi:10.32604/cmc.2021.018486.