Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EXPLAINABLE DEEP LEARNING FRAMEWORK FOR MULTI-CLASS BONE CANCER HISTOPATHOLOGICAL IMAGE CLASSIFICATION USING GRID-OPTIMIZED EFFICIENTNET BONE

GUNTOJU KALPANA DEVI*[1], MAHESH BABU ARRAMA[2]

[18]Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad -500075, Telangana, India

[2]Professor, Department Of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad -500075, Telangana, India.

E-mail: [1*]gkalapana15@gmail.com, [2]maheshbabu.a@klh.edu.in

ABSTRACT

Bone cancer is a rare but aggressive malignancy that primarily affects the skeletal system and poses significant challenges for accurate diagnosis and treatment planning. Among primary bone cancers, osteosarcoma is the most prevalent subtype, requiring precise histopathological evaluation to distinguish between healthy tissue, viable tumor regions, and necrotic tumor areas. However, manual examination of histopathological images is labour-intensive, time-consuming, and prone to inter-observer variability. This study therefore addresses the research problem of whether a lightweight transfer-learning model, when systematically tuned and made explainable, can classify osteosarcoma tissue accurately, stably, and transparently. To address these challenges, this study proposes GO-EfficientNetBone, a Grid-Optimized EfficientNetBone framework for automated multi-class bone cancer tissue classification. The proposed methodology integrates image preprocessing, data augmentation, EfficientNetB0-based deep feature extraction, Global Average Pooling, Grid Search hyperparameter optimization, stratified 10-fold cross-validation, and Grad-CAM explainability within a unified architecture. Histopathological images obtained from the TCIA Osteosarcoma Tumor Assessment Dataset were categorized into Non-Tumor, Viable Tumor, and Necrotic Tumor classes. Experimental evaluation demonstrated that the proposed framework achieved an overall accuracy of 99.04%, precision of 98.97%, recall of 98.91%, F1-score of 98.94%, MCC of 0.985, and a macro-average AUC of 0.998. Furthermore, Grad-CAM localization achieved an average accuracy of 97.33%, providing visual explanations that enhance model transparency and clinical interpretability. Comparative analysis and ablation studies confirmed the superiority of the proposed framework over conventional machine learning and deep learning approaches. The results indicate that GO-EfficientNetBone provides an accurate, robust, and explainable solution for computer-aided bone cancer diagnosis, offering significant potential for supporting pathologists in clinical decision-making and treatment response assessment. The study is limited to a single public dataset analysed at the patch level; future work will therefore focus on patient-level, multi-centre, and whole-slide validation.

Keywords: *Bone Cancer, Osteosarcoma, Histopathological Image Classification, Deep Learning, EfficientNetB0, Transfer Learning, Grid Search Optimization, Explainable Artificial Intelligence, Grad-CAM, Computer-Aided Diagnosis.*

1. INTRODUCTION

Bone cancer is a rare kind of cancer. It begins in the cells of the bone. As a result, it affects the bone structure and its functioning [1]. Bone tumors are classified as either benign or malignant. Malignant tumors are potentially life-threatening due to their uncontrolled growth and ability to spread [2] [3]. Among the first malignant bone tumors, osteosarcoma is the most common and clinically the most important bone tumor, being responsible for approximately 20% of all primary bone tumors [4]. It commonly affects children, adolescents, and young adults during their period of rapid bone growth. Most commonly, the disease appears in long bones of extremities like the femur, tibia, and humerus [5]. It results in immature bone matrix production by malignant osteoblastic cells. Even with improvements in surgical technique and chemotherapy for osteosarcoma, there is still a

high rate of morbidity and mortality, which is especially true if metastasis occurs [6].

Rapid and efficient diagnosis of the first case is highly conducive to a factor boosting treatment outcome. In general, the diagnosis is made through X-ray, CT scan, MRI, biopsy, and histopathological examination [7] [8]. Among these various techniques, histopathological characterization remains the gold standard for diagnosis and response assessment of osteosarcomas. Examination of tumor tissue by pathologists involves assessing vital characteristics like the presence of live tumor cells, necrotic parts of the tumor, and cell and nuclear morphology [9]. They also examine the overall architecture of the tumor tissue. The ratio of necrotic to viable tumor tissue after chemotherapies is especially valuable for therapeutic effectiveness and prognosis prediction. Yet, manual histopathological assessment is a time-consuming, labour-intensive procedure that is prone to inter-observer variability.

Some recent developments in AI, ML, and DL make it possible to extract complex features from medical images autonomously [10]. They are also responsible for classifying these diseases accurately. Deep learning models, especially the Convolutional Neural Network (CNN), have shown outstanding performance in different healthcare applications like cancer detection, tissue segmentation, disease grading, and classification of histopathological images [11]. The adoption of transfer learning models like ResNet, DenseNet, InceptionNet, EfficientNet, etc., will boost the classification accuracy as it draws knowledge from large datasets and applies it to identified medical imaging categories [12]. As a consequence, these techniques lessen dependence on hand-made features and offer powerful ways to learn discriminative tissue representations directly from image data.

According to several studies, experts use AI-aided bone cancer detection with the use of radiographic images, CT scan images, MRI data, radiomics features, and histopathological images [13]. Although these methods are very effective, the existing methods mostly concentrate on only tumour detection, prognosis prediction, or binary classification tasks. In addition, a lot of deep learning models require heavy calculations, require hand-picked hyperparameters, and are black boxes that are hard to interpret [14]. Lack of clinician confidence prevents AI systems from becoming embedded in everyday clinical workflows. Thus, the development of accurate, computable, and interpretable diagnostic models is an important research challenge in osteosarcoma.

To address these challenges, this research introduces a grid-optimized EfficientNetBone (GO-EfficientNetBone) framework to classify osteosarcoma histopathological images. The framework integrates EfficientNetB0-based deep feature extraction, Global Average Pooling, Grid Search hyperparameter optimization, stratified 10-fold cross-validation, and Grad-CAM explainability. The histopathological images of the TCIA Osteosarcoma Tumor Assessment Dataset are classified into Non-Tumor, Viable Tumor, and Necrotic Tumor categories. A comprehensive and clinically interpretable classification framework is proposed to offer accurate and robust classification results for use in diagnostics, treatment response assessment, and decision-making support to pathology experts, through the integration of optimized deep learning and explainable artificial intelligence.

Findings from the experiments indicate that the proposed framework achieved improved classification performance and can effectively be deployed in a computer-aided diagnostic pathology system for bone cancer.

1.1 Research Questions And Objectives

Based on the above discussion, this study addresses the following research problem: can a lightweight transfer-learning model, when its hyperparameters are systematically optimized and its decisions are made visually explainable, classify osteosarcoma histopathology images into Non-Tumor, Viable Tumor, and Necrotic Tumor classes accurately, stably, and transparently? To answer this problem, three research questions (RQs) are formulated.

RQ1: To what extent can an ImageNet pre-trained EfficientNetB0 backbone, combined with Global Average Pooling and a compact classification head, discriminate the three osteosarcoma tissue classes more accurately than conventional machine learning and deep learning baselines?

RQ2: Does exhaustive Grid Search hyperparameter optimization, assessed with stratified 10-fold cross-validation, provide a measurable and stable performance gain over a non-optimized EfficientNetB0 model?

RQ3: Can Grad-CAM provide class-discriminative visual explanations that reliably localize diagnostically relevant tissue regions, so that the predictions can be verified by pathologists?

Corresponding to these questions, the objectives of the study are: (RO1) to design an integrated pipeline for three-class osteosarcoma tissue classification consisting of preprocessing, augmentation, EfficientNetB0 feature extraction, Global Average Pooling and a dense classification head (RQ1); (RO2) to optimize the key hyperparameters systematically through Grid Search and to verify the stability of the resulting model through stratified 10-fold cross-validation (RQ2); (RO3) to integrate Grad-CAM and quantify its localization accuracy for every tissue class (RQ3); and (RO4) to benchmark the framework against conventional machine learning and deep learning baselines and recently published studies, and to measure the contribution of each component through an ablation study (RQ1 and RQ2).

1.2 Research Contributions

From an information technology research perspective, the contribution of this work is primarily a design-oriented artefact contribution, namely an integrated, optimized and explainable classification framework, supported by an empirical evaluation contribution. The work does not propose a new convolutional topology; its novelty lies in the systematic integration, tuning and validation of established components for a clinically relevant task that has received limited attention. The specific contributions are as follows:

(i) A three-class (NT, VT, NCT) histopathology classification framework, GO-EfficientNetBone, that relies only on routinely available H&E-stained images and a lightweight backbone, in contrast to the radiograph-based, prognostic and multimodal approaches reported in the literature.

(ii) An exhaustive Grid Search over five hyperparameters (243 configurations), replacing the manually chosen or narrowly tuned settings reported in earlier studies.

(iii) A rigorous evaluation protocol based on stratified 10-fold cross-validation, multiple complementary metrics (accuracy, precision, recall, F1-score, specificity, MCC and ROC-AUC), comparison with six baselines and an ablation study.

(iv) A quantitative assessment of Grad-CAM explanations through class-wise localization accuracy, which goes beyond the purely qualitative heatmaps commonly reported.

The remainder of this paper is organized as follows. Section 2 critically reviews related work and derives the research gap. Section 3 describes the proposed framework and the complete research method protocol. Section 4 presents the results, the discussion, and the differences from prior research. Section 5 concludes the paper by answering the research questions and outlining the limitations, threats to validity, and future research directions.

2. RELATED WORK

Artificial intelligence (AI), machine learning (ML), deep learning (DL), and radiomics will revolutionize the diagnosis, prognosis, and management of osteosarcoma and other bone tumours [15]. According to traditional diagnostic procedures, the radiological evaluation, its histopathological analysis, and clinical judgment are largely subjective and less reproducible. Researchers are increasingly giving priority to AI frameworks that offer automated and accurate clinical diagnostic support systems.

Wang et al. (2026) [16] investigated the AI-based prognostic risk stratification for osteosarcoma and bone metastatic cancer. A study demonstrated how some artificial intelligence algorithms can analyze clinical and pathological data to predict patient outcomes and provide suggestions for personalized treatments. AI-assisted prognosis prediction can help improve the management of patients, as shown in the framework. Nonetheless, they were more concerned with the prognosis of survival and risk than with automated tissue classification.

Kiani Kalejahi et al. (2026) [17] introduced a lightweight deep learning architecture called Bone-CNN for the multi-class classification of primary bone tumors based on radiographic images. The framework used a computationally efficient convolutional neural network capable of classifying various categories of bone tumors with low computational complexity. Results from the study show that it has vast classification potential and can perform well in clinical settings. Nonetheless, the model was developed with the

aid of radiographic images, and it did not assess histopathological classification of tissues.

Alharran et al. (2026) [18] generated algorithms and a web-based prognostic tool for machine learning for different histological subtypes of osteosarcoma. Through a retrospective cohort study using clinical variables and machine learning models, their study was able to predict patient outcomes and assist clinical decision-making. The prognostic system proposed could be useful for predicting the survival of Osteosarcoma, but there is no image-based tumor classification or explainable AI.

Bulut et al. (2026) [19] provided a complete review of the applications of artificial intelligence and machine learning in bone metastasis management. According to the authors, recent developments in tumour detection, disease monitoring, treatment planning and outcome predictions through AI-based approaches. In a review on machine learning in orthopedic oncology, the authors identify data heterogeneity, interpretability and integration in the clinical workflow as major issues. Even though the work does provide some information, it is mostly a survey and does not offer any classification scheme.

According to Song et al. (2026) [20], the authors first designed a deep learning-based multimodal fusion framework paired with imaging, pathology, and circulating tumor cell (CTC) information for the early diagnosis of pediatric distal femur osteosarcoma. The approach was shown to provide better diagnostic sensitivity and classification accuracy than single-modality systems by making use of heterogeneous data sources. The study showed that multimodal learning can obtain complementary diagnostic information. Nonetheless, the demand for various data modalities raises complications of implementation and could impede actual deployment in resource-poor health settings.

Sergioli et al. (2026) [21] proposed a multi-class quantum-inspired classification framework for radiomics cancer classification. Even if their work is focused on lung cancer subtyping and prostate cancer risk stratification, the classifier showed that the application of high-level computational intelligence techniques on high-dimensional radiomic features is possible. The results show that classification performance is improved and the ability to discriminate features is enhanced. However, the framework was not developed for the classification of osteosarcoma histopathology images.

According to one of the latest osteosarcoma-specific studies, Borji et al. (2025) [22] developed a hybrid deep learning framework for osteosarcoma histopathology image analysis. The model combined convolutional neural networks and transformer models to get local texture information and global contextual understanding. Experimental results showed superior classification performance relative to traditional deep learning techniques. Though this framework is successful, it suffers from increased complexity and lacks an explainability mechanism for the interpretability of the classifications.

Li et al. (2025) [23] demonstrated the development of a multimodal diagnostic framework combining hyperspectral imaging and deep learning for differentiating osteosarcoma and bone callus tissue. When spectral data is coupled with spatial data, tissue characterization accuracy can improve, which will enhance diagnostics. Even though the proposed configuration gave better results, the use of proprietary hyperspectral imaging equipment may reduce the clinical applicability.

Rao and Madhavi (2025) [24] designed a framework using explainable AI to identify malignant bone tumours in the Osteosarcoma Tumour Assessment dataset. Their method, which included explainability, used optimized pre-trained deep learning models for better diagnosis. The research states that explainable artificial intelligence can enhance physicians' trust as well as help them understand the predictions of automations. Nonetheless, hyperparameter optimization was limited to just two values. A more complete hyperparameter optimization may yield a performance boost.

Brar et al. (2025) [25] summarize the recent advances in osteosarcoma, which include its molecular biology, targeted therapies, immunotherapy, and AI-assisted diagnosis. According to the authors, the use of AI systems may enhance disease detection and treatment planning. The review also traces the therapeutic management of osteosarcoma from standard-of-care protocols to precision therapeutics. As per the review, computational intelligence and personalized treatment are becoming a crucial aspect in oncology. However, as a narrative

review, it does not propose or experimentally evaluate any computational diagnostic model.

Methods based on radiomics have also acquired much attention recently. Wang and colleagues (2025) [26] set up a radiomics framework based on different machine learning algorithms to identify early bone metastases that were negative on CT. The findings indicate that radiomic feature extraction can identify obscure pathological features that cannot be easily identified by eyeballing. Bai et al. (2025) [27] proposed the framework of multi-class segmentation and deep radiomics for multi-centre lumbar spine MRI. They proposed wavelet fusion-based two-stage classifiers with explainable machine learning to improve classification accuracy. Nevertheless, both studies operate on radiological images using radiomic pipelines and therefore cannot capture the cellular-level differences between viable and necrotic tumour that histopathology reveals.

Liang et al. (2025) [28] researched the use of AI technology in assessments and imaging of bone tumours. Researchers have highlighted the effectiveness of deep learning algorithms in enhancing sensitivity and slashing diagnostic delay. A multimodal microscopic imaging framework integrated with deep learning techniques for osteosarcoma diagnosis was proposed (Wang et al., 2025) [29]. Use of multiple imaging modalities allowed for more comprehensive tissue characterization and improved classification performance. However, [28] concentrates on radiological imaging of older adults, and [29] depends on multimodal microscopic acquisition that is not part of routine H&E-based pathology workflows, which restricts its practical adoption.

According to Wang et al. (2025) [30], spinal osteosarcoma and chondrosarcoma can be differentiated using a model based on radiomics trained using preoperative multi-modal imaging data. The study reported the role of radiomic features derived from different imaging modalities in improving diagnostic discrimination ability. Dalai et al. (2025) [31] suggested an automated framework for the detection of bone cancer with the help of deep learning and the use of X-ray images, which provided encouraging results while reducing reliance on the manual interpretation of results. However, [30] addresses differential diagnosis between tumour types rather than intra-tumour tissue states, and the X-ray-based detector in [31] cannot quantify chemotherapy-induced necrosis, which is the key histological indicator of treatment response.

Studies based on radiomics have investigated the use of deep learning and machine learning in classifying bone diseases. Researchers Liu and colleagues (2025) [32] created a deep learning radiomics model using computed tomography images to forecast the category of osteoporotic vertebral fractures. Mohammadi-Sadr et al. (2025) [33] developed a model that provides bone mineral density (BMD) from routine computed tomography (CT) scans through the use of radiomics, bone morphometry, and Hounsfield unit-derived features. The studies cited above collectively show the increasing role of radiomics and deep learning in orthopedic imaging and diagnostic evaluation. These studies, however, target osteoporotic fractures and bone mineral density rather than malignancy, so their relevance to osteosarcoma histopathology is limited to general methodological insights.

Earlier contributions have laid an important foundation for AI-assisted bone cancer diagnosis. Sampath et al. (2024) [34] carried out a comparative analysis of various CNN architectures for early bone cancer diagnosis using CT images. The findings highlighted that the network architecture selection affects classification performance. Wen et al. (2024) [35] used WGCNA and machine learning to establish hypoxia-related gene signatures for diagnosing and predicting the prognosis of osteosarcoma. The potential of molecular biomarkers and machine learning for predicting diseases is demonstrated by the proposed framework. However, [34] focuses on benchmarking architectures on CT images rather than on an optimized and explainable tissue-level pipeline, and [35] relies on gene-expression data that require molecular assays instead of routinely available tissue slides.

Abedi et al. (2024) [36] studied miRNA expression models based on machine learning for metastatic osteosarcoma. The researchers discovered key molecular signatures that can distinguish whether a case is metastatic or non-metastatic. Greenwood et al. (2024) [37] published Osteosarcoma: A Clinical Overview of the disease, its diagnosis, and treatment. The study isn't about AI techniques, but it is valuable for any diagnostic computational research in clinical contexts. Similar to [35], the approach of [36] depends on expression profiling, which limits its use in routine diagnostic settings,

whereas [37] provides useful clinical context but no computational method.

A new three-class annotation approach for AI-based detection of early osteosarcoma on conventional radiographs is proposed by Hasei et al. (2024) [38]. The research showed that improved annotation methods can significantly enhance model performance for rare cancer diagnosis. Zhang et al. (2024) [39] used their methodology on X-ray images to classify osteoporotic vertebral fractures. They noted the effectiveness of classifying with an integrated radiomic feature extraction and deep learning classifier. Both studies, however, operate on plain radiographs, in which cellular morphology and necrosis cannot be observed, and therefore cannot support histological assessment of treatment response.

2.1 Critical Analysis Of Existing Studies

A critical reading of the studies reviewed above shows that, although AI-assisted bone cancer analysis has advanced rapidly, the specific need addressed in this work remains largely unmet. Table 1 summarizes representative studies in terms of the data used, the task addressed, and their principal limitation with respect to automated, explainable, three-class histopathological tissue classification.

Table 1: Critical Summary Of Representative Prior Studies

Study	Data / Modality	Task Addressed	Principal Limitation Relative To This Work
Wang et al. [16]	Clinical and pathological variables	Prognostic risk stratification	No image-based tissue classification
Alharran et al. [18]	Clinical variables (retrospective cohort)	Survival prediction and web-based tool	No image analysis and no visual explanation
Kiani Kalejahi et al. [17]	Radiographs	Multi-class bone tumour classification	No cellular-level information; necrosis cannot be assessed
Dalai et al. [31]	X-ray images	Bone cancer detection	Detection-level output without pathological detail
Hasei et al. [38]	Plain radiographs	Early osteosarcoma detection	Annotation-focused; no histological response assessment
Wang et al. [30]; Bai et al. [27]	CT / MRI radiomics	Differential diagnosis; spine segmentation	Radiological features only; not applicable to tissue slides
Song et al. [20]	Imaging, pathology and CTC data	Multimodal early diagnosis	Several modalities required; high acquisition and computational cost
Li et al. [23]	Hyperspectral images	Osteosarcoma versus bone callus	Requires specialized hyperspectral equipment
Wang et al. [29]	Multimodal microscopic images	Osteosarcoma diagnosis	Depends on non-routine imaging modalities
Sampath et al. [34]	CT images	Comparison of CNN architectures	Benchmarking only; no integrated, optimized, and explainable pipeline
Wen et al. [35]; Abedi et al. [36]	Gene / miRNA expression	Diagnosis, prognosis, metastasis prediction	Requires molecular assays; not image-based
Borji et al. [22]	Histopathology images	Hybrid CNN-Transformer classification	High model complexity; no explainability mechanism
Rao and Madhavi [24]	Histopathology images (TCIA)	Bone cancer detection with explainable AI	Hyperparameter optimization limited to two values
Bulut et al. [19]; Brar et al. [25]	Published literature	Narrative reviews	No classification model proposed or evaluated

Four recurring weaknesses emerge from this analysis. First, a substantial part of the literature [16], [18], [35], [36] targets prognosis or molecular prediction rather than image-based tissue classification and therefore does not support histological treatment-response assessment, which depends on distinguishing viable from necrotic tumour. Second, most image-based models rely on radiographs, CT or MRI [17], [30], [31], [38], which lack the cellular-level evidence required for necrosis

estimation, while multimodal and hyperspectral systems [20], [23], [29] improve accuracy at the cost of acquisition protocols that are rarely available in routine pathology laboratories. Third, among histopathology-based approaches, the hybrid CNN-Transformer model [22] increases computational demand and provides no explanation mechanism, whereas the explainable approach in [24] explores only a narrow hyperparameter range, so its reported performance may be sub-optimal and difficult to reproduce. Fourth, the reviewed studies seldom combine systematic hyperparameter search, stratified cross-validation with fold-wise variance, and quantitative evaluation of visual explanations within a single framework.

These weaknesses define the need for the present work: a computationally light framework that uses only routine histopathology images, is systematically tuned, is rigorously validated, and is quantitatively explainable for the three-class classification of osteosarcoma tissue. The problem statement and the research gap derived from this analysis are formalized in Sections 2.2 and 2.3.

2.2 Problem Statement

Bone cancer or osteosarcoma is a malignant tumour that can be very aggressive and requires a proper histopathological diagnosis. The manual checking of histopathological images is laborious, time-consuming, and subject to inter-observer variability that can lead to discrepancies in diagnosis. The classification performance of automated bone cancer detection is still impacted by the presence of issues such as high inter-class similarity of types of tissue, low interpretability of the model, under-optimized hyperparameters, and low generalization capability. As a consequence, there is a need for the development of an efficient, robust, explainable deep learning framework that can classify Non-Tumor, Viable Tumor, and Necrotic Tumor tissues with high accuracy from histopathological images along with reliable visual explanations.

2.3 Research Gap

Recent studies have successfully used AI, deep learning, radiomics, and multimodal learning techniques for the diagnosis and prognosis of bone cancer [40]. But mostly, the existing approaches only work towards radiography imaging, prognosis prediction, or binary classification problems, and similar tasks. Moreover, many deep learning frameworks are dependent on manually selected hyperparameters, do not use explainable AI, and are black-box models that limit their clinical acceptance [41]. Also, the implementation of such models may involve high computational complexity, which makes them difficult to deploy. Another limitation is the lack of sufficient validation strategies that are rigorous, for example, stratified cross-validation. As a result, there is still a requirement for a computationally efficient and clinically interpretable framework comprising optimized deep feature extraction, hyperparameter tuning, validation, and visual explainability for accurate classification of bone cancer tissue.

The GO-EfficientNetBone was proposed to overcome these issues by utilizing EfficientNetB0-based feature extraction, Grid Search hyperparameter optimization, stratified 10-fold cross-validation, and Grad-CAM explainability, all within the same architecture for automatic multi-class classification of labelled bone cancer histopathological images. Accordingly, the research questions formulated in Section 1.1 follow directly from these gaps: RQ1 addresses accurate tissue-level classification, RQ2 addresses hyperparameter optimization and rigorous validation, and RQ3 addresses explainability.

3. METHODOLOGY

This section introduces the GO-EfficientNetBone framework for automatic multi-class bone cancer classification. The method includes dataset acquisition, image preprocessing, deep feature extraction using EfficientNetB0, Global Average Pooling, hyperparameter optimization using Grid Search, multi-class classification, stratified 10-fold cross-validation, Grad-CAM explainability, and performance evaluation. To improve data quality and generalization, histopathological images undergo preprocessing steps like resizing, normalization, and augmentation. EfficientNetB0 was used for feature extraction, and Grid Search was used to optimize the hyperparameters of the classification network. Ultimately, the Grad-CAM provides a visual-based explanation or reasoning of the classification decisions, while extensive evaluation metrics assess the effectiveness, robustness, and clinical applicability of the proposed framework.

3.1 Overview Of The Proposed Framework

The proposed study combined the EfficientNetBone framework with grid optimization for the automated osteosarcoma histopathological image classification for three classes. The framework combines deep transfer learning, hyperparameter optimization, explainable artificial intelligence, and rigorous validation to accurately distinguish between Non-Tumor (NT), Viable Tumor (VT), and Necrotic Tumor (NCT) tissue regions. The complete Architecture is illustrated in Figure 1.

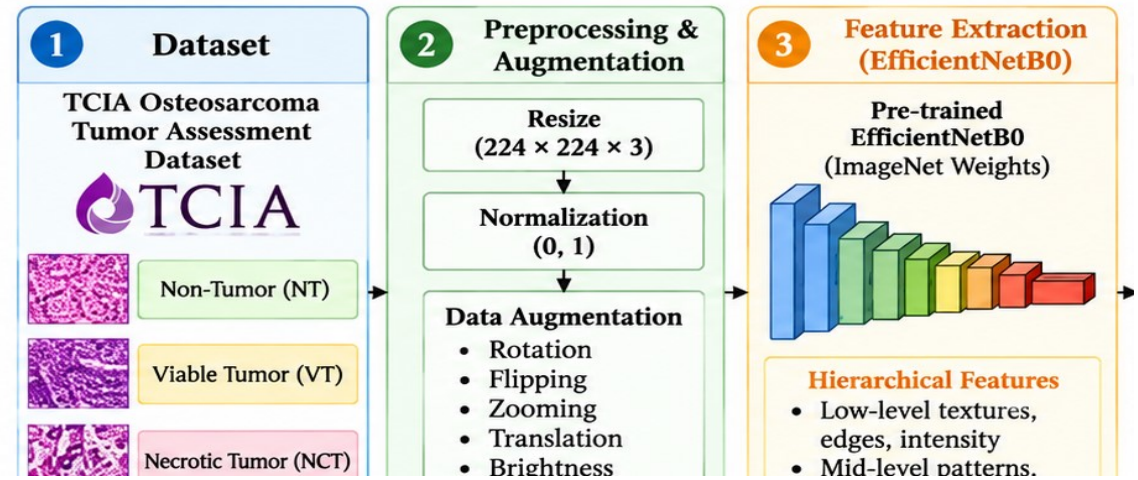


Figure 1: Architecture Of Grid-Optimized EfficientNetBone (GO-EfficientNetBone) Framework

The proposed methodology consists of the following sequential stages:

3.1.1 Dataset acquisition

The basis of this work is the TCIA Osteosarcoma Tumor Assessment dataset. This dataset contains histopathological image patches of three types: Non-Tumor (NT), Viable Tumor (VT), and Necrotic Tumor (NCT). The annotated tissue images are the input data. Thus, the model training and evaluation are done using this data.

3.1.2 Image preprocessing and augmentation

To ensure that all images have the same input quality of 224 × 224 × 3, it is resized and normalized within the [0,1] range. Data augmentation techniques applied are rotation, flipping, zooming, translation, and brightness to increase data variation and improve model generalization.

3.1.3 Deep feature extraction (EfficientNetB0)

An EfficientNetB0 that was pre-trained with ImageNet weights automatically extracts hierarchical image features. The network determines low-level textures, mid-level forms, and high-level tumor properties for classifying tissues.

3.1.4 Global average pooling (GAP)

Global Average Pooling processes the high-dimensional feature maps generated by EfficientNetB0 into compact feature vectors. This process cuts down the trainable parameters while retaining essential discriminative information.

3.1.5 Classification head

The feature vector is applied to dense layers with 256 neurons, ReLU activation, and a dropout layer with 0.3. A final prediction category is then obtained using a Softmax output layer that produces probabilities of the three tissue classes.

3.1.6 Grid search hyperparameter optimization

Grid Search is a process for systematically testing multiple combinations of learning rate, batch size, dropout rate, dense units, and epoch. The best configuration is chosen based on validation accuracy for enhancing classification performance and model stability.

3.1.7 Training and multi-class classification

The network is trained using the optimal hyperparameter configuration identified by Grid Search. During the training phase, the neural network learns the discriminative tissue representation from the training dataset and multi-classifies the histopathological image of osteosarcoma into NT, VT, and NCT in each fold.

3.1.8 Performance evaluation and explainability

The precision, recall, F1-Score, specificity, the MCC, and the ROC-AUC are used to measure model performance. In addition, Grad-CAM

explainability generates heatmaps that highlight the diagnostically relevant tissue regions, allowing the model's classification decisions to be visualized and verified.

The GO-EfficientNetBone framework achieves high accuracy and clinical interpretability to classify these tissue types by integrating preprocessing, deep feature extraction using EfficientNetB0, optimization using grid search, multi-class classification, and explainability using Grad-CAM into a single pipeline.

3.2 Dataset Description

The Cancer Imaging Archive (TCIA) provides the public Osteosarcoma Tumor Assessment Dataset [42], which we used to evaluate the proposed GO-EfficientNetBone framework. The dataset consists of histopathological image patches that were collected from osteosarcoma patients and annotated by expert pathologists into three clinically relevant classes, namely, Non-Tumor (NT), Viable Tumor (VT), and Necrotic Tumor (NCT). These were applied for assessing tumor response after chemotherapy and are useful in treatment evaluation.

The dataset is histopathological image patches extracted from osteosarcoma tissue samples from patients receiving neoadjuvant chemotherapy. Expert diagnosticians labelled every slice into one of three tissue types of clinical interest. The Classes of the dataset are given in Table 2.

Table 2: Dataset Classes

Class	Description
Non-Tumor (NT)	Healthy tissue regions without malignant cells
Viable Tumor (VT)	Living osteosarcoma cells indicate active tumor growth
Necrotic Tumor (NCT)	Dead tumor tissues resulting from the treatment response

3.2.1 Attributes used in the study

The study deals with the histopathological image dataset, which is more of a visual characteristic than a table attribute. The tissue features automatically learned by our system fall into three categories, namely Morphological (nuclear shape and size, cell density, arrangement and boundary), Texture (tissue granularity, chromatin distribution, staining patterns, necrotic appearance, and heterogeneity), and Spatial (tumour organization, tissue architecture, stromal arrangement, invasion and necrosis). These attributes are obtained through deep convolutional feature learning.

3.3 Image Preprocessing

To minimize the noise generated due to the slight variations in staining, scanner settings, illumination, and other conditions, a comprehensive preprocessing pipeline was deployed before the model training step. To meet the input requirements of EfficientNetB0 and ensure that all images underwent the same feature extraction, all images were resized to 224 × 224 × 3 pixels. For normalization of the pixel intensity, we scale down the pixel value of the image into the range of 0 to 1. Thus, enhancing numerical stability and accelerates the convergence of the model.

To improve the robustness of the model against overfitting, we also apply some data augmentation techniques during training, which include random rotation, horizontal and vertical flipping, zooming, shifting, and brightness change. The transformations enrich the dataset diversity and set the network to learn invariant tissue representation under varied imaging conditions. Upon executing this preprocessing stage, it enhances the generalization ability, which ultimately improves the proposed framework's classification performance.

3.4 Deep Feature Extraction Using EfficientNetB0

The importance of deep feature extraction in the GO-EfficientNetBone framework is to automatically learn the discriminative tissue features of osteosarcoma histopathological images. For its classification accuracy and level of effort, EfficientNetB0 was chosen as the backbone convolutional neural network in this study. The EfficientNetB0 does not use a standard convolutional network like others. It uses something called compound scaling. This compound scaling scales almost uniformly the depth, width, as well as image resolution of the model. This results in the learning of enhanced representations. Also, all this happens while maintaining a low overall cost.

A pre-trained EfficientNetB0 model initialized with ImageNet weights was adopted through transfer learning. This approach enables the network to leverage previously learned visual knowledge and adapt it to the domain of histopathological image analysis. Given an input image X, the EfficientNetB0 network generates a deep feature representation F, expressed as in Eq. (1)

$$F = EfficientNetB0(X) \quad (1)$$

where X denotes the input histopathological image, and F represents the extracted deep feature tensor.

The network develops different layers of meaning from its data. The first few layers capture low-level features such as colour distributions, nuclear intensity, cellular boundaries, and texture patterns. Intermediate layers detect cellular arrangements, tumor, stroma, and necrosis. The deeper layers of the network can provide high-level semantic information that includes viable tumor regions, necrotic tissue areas, tumor heterogeneity, and tissue organization patterns. The resulting learned representations form the basis for accurate and robust multi-class tissue classification.

3.5 Feature Aggregation Using Global Average Pooling

The EfficientNetB0's feature maps contain numerous spatial and semantic information in its manifold channels. Global Average Pooling (GAP) is used to reduce these complex feature maps to a descriptor. GAP produces a single value after averaging the activations of the feature map over its spatial dimensions. Thus, it is a way to generate a fixed-length feature vector from the feature maps obtained.

The GAP operation is mathematically defined as in Eq. (2)

$$G_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{ijc} \quad (2)$$

where H and W denote the spatial dimensions of the feature map, F_{ijc} represents the activation value at the location (i, j) of channel c, and G_c denotes the aggregated feature value for that channel.

GAP reduces the number of trainable parameters to a great extent as well as the chances of overfitting, compared with traditional fully connected flattening operations. Furthermore, it retains the most useful data gathered by the convolutional layers while enhancing the model's generalization power. As a result, GAP helps in bridging feature extraction and classification.

3.6 Classification Network

After feature aggregation, the compact feature vector is passed to a fully connected classification network that classifies each histopathological image into one of the three tissue classes. Adaptation of neural networks is done with the help of classification networks. Then they use a dense layer. Regularization by dropout takes place. Later, a softmax output layer completes the process.

The aggregated feature vector G is first passed through a dense layer containing 256 neurons with Rectified Linear Unit (ReLU) activation in Eq. (3)

$$H = ReLU(WG + b) \quad (3)$$

where W represents the learnable weight matrix and b denotes the bias vector.

The dense layer learns nonlinear relationships among extracted features and enhances class separability. To improve model generalization and reduce overfitting, a dropout layer with a probability of 0.3 is subsequently applied in Eq. (4)

$$D = Dropout(H, 0.3) \quad (4)$$

The dropout mechanism randomly deactivates a subset of neurons during training, thereby preventing excessive dependence on specific feature representations.

Finally, a Softmax output layer consisting of three neurons generates class probabilities corresponding to Non-Tumor, Viable Tumor, and Necrotic Tumor tissues:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}} \quad (5)$$

The class that is predicted is the one with the maximum probability value. An effective classification network converts deep features to tissue-level predictions.

3.7 Grid Search Hyperparameter Optimization

The selection of suitable hyperparameters plays a critical role in the performance of deep learning models. Consequently, an exhaustive Grid Search optimization strategy was adopted to ascertain the most favourable parameters for the classification of osteosarcoma tissues. Grid Search examines every possible combination of hyperparameters in a given space and selects one with the highest validation accuracy.

The search space included learning rates, batch sizes, dense layer units, dropout rates, and training epochs. Specifically, learning rates of 0.01, 0.001, and 0.0001; batch sizes of 16, 32, and 64; dense layer sizes of 128, 256, and 512 units;

dropout rates of 0.2, 0.3, and 0.5; and epoch values of 30, 50, and 100 were investigated.

The optimal hyperparameter set was selected according to Eq. (6)

$$\theta^* = \arg\max Accuracy(\theta) \quad (6)$$

where θ denotes a candidate hyperparameter configuration and θ^* represents the optimal parameter set.

Following experimental analysis, the best-performing architecture was determined to be one with a learning rate of 0.0001, batch size of 32, dense layer of 256 units, dropout rate of 0.3, optimizer of Adam, and 50 training epochs. The enhancement of convergence stability and reduction in classification errors due to these optimized settings were significant contributory factors to the performance of the proposed GO-EfficientNetBone design.

3.8 Grad-CAM Explainability

To effectively use artificial intelligence systems for medical diagnosis, they must be capable of being interpreted. To improve model transparency and visual explanation of its classification decisions, Gradient-weighted Class Activation Mapping (Grad-CAM) was added to the proposed framework.

Unlike many existing methods, Grad-CAM generates class-discriminative heatmaps by locating the regions of the image that are most relevant for the prediction of interest. The process determines the flow of gradients originating from the target class that culminates within the last layer of convolution. Based on this, the importance of every feature map is estimated.

The Grad-CAM importance weights are computed as in Eq. (7)

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (7)$$

where y^c represents the score of class c, A_{ij}^k denotes the activation value of feature map k, and Z represents a normalization factor.

The final localization map is generated as in Eq. (8)

$$L_{GradCAM} = ReLU(\sum_k \alpha_k A^k) \quad (8)$$

The heatmaps produced indicate the presence of viable tumor cells, necrotic areas, and healthy tissue structures important for diagnosis. With the help of these visual explanations, pathologists can verify model predictions. They also help in increasing confidence in the automated classification process.

3.9 Performance Evaluation Metrics

To evaluate the performance of the proposed GO-EfficientNetBone framework, various quantitative metrics were deployed. The classifications were evaluated for accuracy, robustness, reliability, and generalization across all tissues, and measured accuracy, precision, recall, F1-score, specificity, and Matthews correlation coefficient (MCC). Accuracy measures how often a classifier is correct overall, while Precision measures how often a classifier's positive prediction is correct. The recall is indicative of the ability of the model to identify all types of relevant samples. The F1-Score gives an average of both the Precision and the Recall. Specificity tests the classifier's ability to send the negative samples to the negative class. The Matthews correlation coefficient (MCC) is a more robust measure of quality, especially if used on data with an imbalanced distribution of classes.

For each tissue category and the macro-average classification performance, Receiver Operating Characteristic (ROC) analysis and Area Under the Curve (AUC) measurements were performed to assess class separability. Also, mean classification accuracy with standard deviation over ten-fold cross-validation was performed for statistical validation. These metrics offer an assessment of the proposed framework and enable comparison with most competing machine learning and deep learning methods.

3.10 Computational Environment

The experiments were conducted in a high-performance computing environment that was implemented based on medical image analysis using deep learning. The TensorFlow/Keras deep learning framework was used to develop the GO-EfficientNetBone, while the implementation is in Python 3.11. Other image pre-processing, augmentation, visualization, and performance evaluation are done with OpenCV, NumPy, Pandas, and Scikit-learn libraries.

The model was trained and inferred using an NVIDIA RTX-series graphics processing unit

(GPU). To carry out computing operations, an Intel Core i7/i9 CPU with 32 GB RAM was utilized. The ability of the environment to use the GPU reduced a lot of training time. Grid Search was used to systematically optimize the hyperparameters. The computational setup was designed to conduct efficient, scalable, reproducible, and reliable assessment of the proposed framework at each stage of the experiment.

3.11 Research Method Protocol

To make the results traceable and reproducible, the complete research method protocol applied to obtain the reported results is described below. The study follows a quantitative experimental research design in which the proposed framework is constructed, optimized, and evaluated against baseline methods under an identical evaluation procedure. The protocol consists of the following eight steps.

Step 1 – Problem formulation: The research problem and research questions RQ1–RQ3 (Section 1.1) were derived from the critical analysis of the literature (Section 2.1). The evaluation criteria were fixed before experimentation: accuracy, precision, recall, F1-score, specificity, MCC, and ROC-AUC for classification, and localization accuracy for explainability.

Step 2 – Data acquisition: The publicly available TCIA Osteosarcoma Tumor Assessment dataset [42] was obtained, and the histopathological image patches carrying one of the three expert-assigned labels (NT, VT, or NCT) were used as the study corpus (Table 2).

Step 3 – Data partitioning: The corpus was divided into ten stratified folds that preserve the class proportions of the dataset. In each of the ten iterations, nine folds were used for training, and the remaining fold was held out for testing, so that every image was tested exactly once. The validation accuracy used for model selection was computed on data drawn from the training folds only, so the held-out test fold was never used for tuning.

Step 4 – Preprocessing and augmentation: All images were resized to 224 × 224 × 3 pixels and normalized to the range [0, 1] (Section 3.3). Rotation, horizontal and vertical flipping, zooming, shifting, and brightness adjustment were applied on-the-fly to the training images only, whereas the test images were only resized and normalized.

Step 5 – Model construction: The ImageNet pre-trained EfficientNetB0 backbone was followed by Global Average Pooling, a dense layer with ReLU activation, a dropout layer, and a three-neuron Softmax output layer (Sections 3.4–3.6). The network was trained with the categorical cross-entropy loss and the Adam optimizer.

Step 6 – Hyperparameter optimization: Grid Search evaluated all $3 \times 3 \times 3 \times 3 \times 3 = 243$ combinations of learning rate {0.01, 0.001, 0.0001}, batch size {16, 32, 64}, dense units {128, 256, 512}, dropout rate {0.2, 0.3, 0.5} and epochs {30, 50, 100}. The configuration with the highest validation accuracy (Eq. 6, Table 3) was fixed and used in all subsequent experiments.

Step 7 – Training, testing and comparison: Using the optimal configuration, the model was trained and tested on each of the ten folds; fold-wise metrics were recorded and reported as mean ± standard deviation, together with the confusion matrix and ROC curves. The baseline methods (SVM, Random Forest, XGBoost, CNN, ResNet50 and EfficientNetB0) were evaluated on the same data partitions, and the ablation variants were obtained by adding the framework components incrementally.

Step 8 – Explainability and comparative analysis: Grad-CAM heatmaps were generated from the last convolutional layer of EfficientNetB0 for the test images, and the class-wise localization accuracy was computed as the percentage of heatmaps whose dominant activation region coincided with the diagnostically relevant tissue of the corresponding class. Finally, the results were compared with recently published studies (Table 11), and the differences, shortcomings, and threats to validity were analysed (Sections 4.10 and 5.2).

4. RESULTS AND DISCUSSION

This section examines the experimental assessment of GO-EfficientNetBone for classifying osteosarcoma histopathological images. We tested our framework on the TCIA osteosarcoma tumour assessment dataset with three tissue types Non-Tumor (NT), Viable Tumor (VT), and Necrotic Tumor (NCT). The experiments were performed with stratified 10-fold cross-validation, which provides a good performance estimate, minimizing sampling bias. The proposed framework involves

EfficientNetB0-based deep feature extraction, grid search hyperparameter tuning, and Grad-CAM explainability for accurate classification with clinical interpretability.

4.1 Hyperparameter Optimization Results

The performance of a deep learning model relies on hyperparameter selection. A systematic Grid Search optimization strategy was used to find a suitable network architecture for osteosarcoma tissue classification. Various learning rates, batch sizes, dropout rates, dense size, and epoch combinations were examined for learning. The optimal hyperparameters identified by the Grid Search process are presented in Table 3.

Table 3: Optimal Hyperparameter Configuration Obtained Using Grid Search

Hyperparameter	Optimal Value
Backbone Network	EfficientNetB0
Learning Rate	0.0001
Batch Size	32
Dense Units	256
Dropout Rate	0.3
Optimizer	Adam
Epochs	50

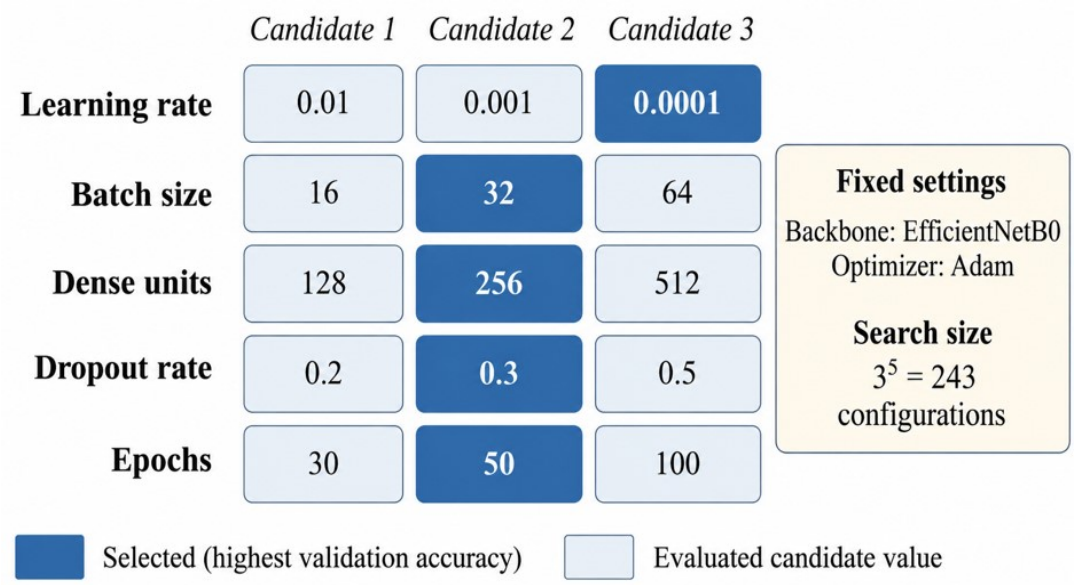


Figure 2: Grid Search Space And Selected Optimal Hyperparameter Configuration For The Proposed GO-EfficientNetBone Framework

The optimal hyperparameters that we found for the proposed GO-EfficientNetBone model through Grid Search optimization are shown in Figure 2. The configuration uses EfficientNetB0 as the backbone network, with a learning rate of 0.0001, a batch size of 32, dense units of 256, dropout of 0.3, Adam as the optimizer, and 50 epochs. The chosen parameters yielded the best results and hence were selected for future experiments. The figure also shows the candidate values evaluated for each hyperparameter, with the selected values highlighted.

4.2 Comparative Classification Performance

The study compares the performance of the proposed framework with various conventional machine learning and deep learning approaches, which are often used for medical image classification. The comparison is made with SVM, RF, XGBoost, vanilla CNN, ResNet50, and standard EfficientNetB0. The comparative performance results are summarized in Table 4.

Table 4: Comparative Performance Analysis Of Different Classification Methods

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MCC
SVM	91.42	90.87	90.33	90.59	0.87
Random Forest	92.68	92.11	91.84	91.97	0.89
XGBoost	94.15	93.92	93.61	93.76	0.91
CNN	95.27	95.11	94.82	94.96	0.93
ResNet50	97.31	97.08	96.94	97.01	0.96
EfficientNetB0	98.12	97.96	97.83	97.89	0.97
GO-EfficientNetBone	99.04	98.97	98.91	98.94	0.985

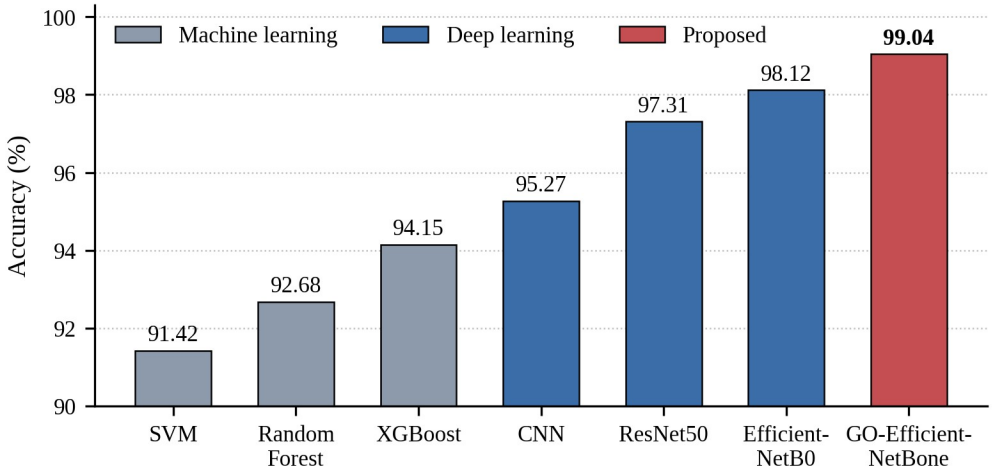


Figure 3: Comparative Classification Accuracy Of Different Methods.

As shown in Figure 3, the GO-EfficientNetBone framework consistently outperforms all other techniques. The accuracy was improved by 0.92% compared with the EfficientNetB0 model. Additionally, the MCC score of 0.985 suggests a remarkable agreement between the predicted and actual labels, confirming the robustness of the classifier discussed in this study.

4.3 Class-Wise Performance Evaluation

A separate performance analysis on a class-wise basis was performed to examine the proposed framework's ability to classify each tissue class individually. The Non-Tumor category had the highest precision and specificity, as healthy tissue regions have very distinct morphological characteristics. The framework shows potential for the identification of chemotherapy-induced necrotic tissue regions, as the largest recall value was observed in the Necrotic Tumor category. The class-wise Precision, Recall, F1-Score, and Specificity values are shown in Table 5.

Table 5: Class-Wise Classification Performance Of The Proposed Framework

Class	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)
Non-Tumor	99.42	99.11	99.26	99.83
Viable Tumor	98.71	98.52	98.61	99.47
Necrotic Tumor	98.79	99.1	98.94	99.52
Average	98.97	98.91	98.94	99.61

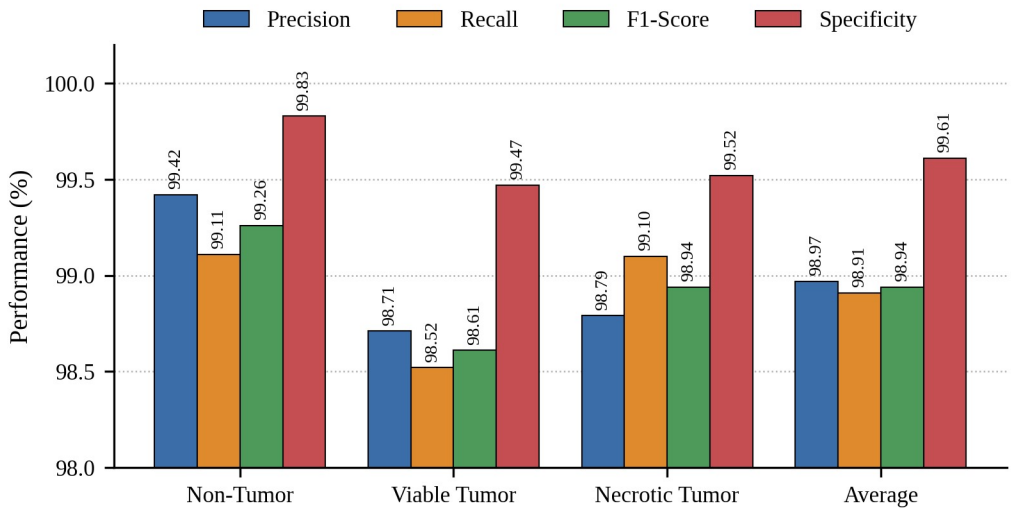


Figure 4: Class-Wise Classification Performance Of The Proposed GO-EfficientNetBone Framework.

Figure 4 presents the precision, Recall, F1-Score, and specificity achieved for non-tumor, viable tumor, necrotic tumor, and average. The proposed framework achieves consistently high performance over all classes and yields more than 98.5% for all evaluation metrics, demonstrating a robust and reliable classification of osteosarcoma tissue. The proposed framework is robust as it performs well in all categories of the evaluation.

4.4 Confusion Matrix Analysis

The confusion matrix provides information about the samples classified correctly by the model and incorrectly, along with the classes they belong to. It helps in understanding various patterns of confusion in classes and the reliability of classifications. The confusion matrix obtained using the proposed GO-EfficientNetBone framework is given in Table 6.

Table 6: Confusion Matrix Of The Proposed Framework

Actual / Predicted	Non-Tumor	Viable Tumor	Necrotic Tumor
Non-Tumor	594	2	4
Viable Tumor	3	589	8
Necrotic Tumor	2	5	593

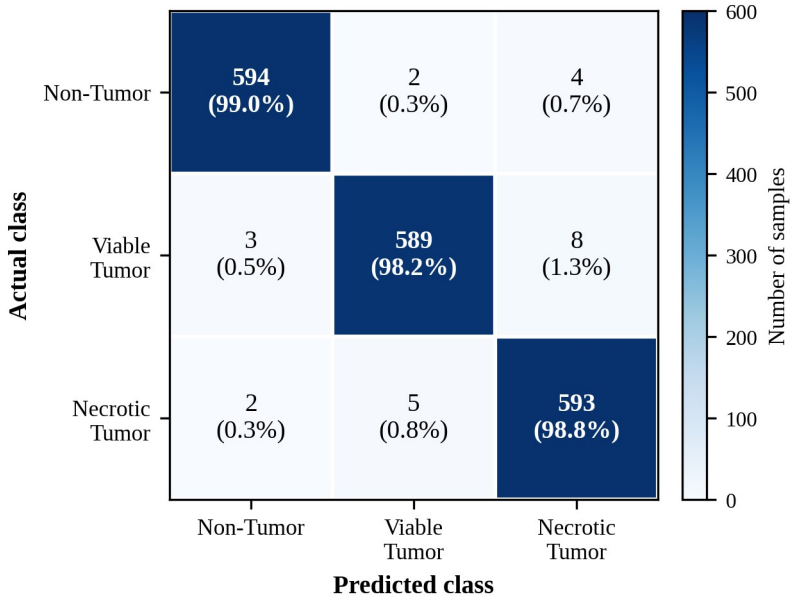


Figure 5: Confusion Matrix Of The Proposed GO-EfficientNetBone Framework.

As seen in Figure 5, the confusion matrix shows the classification performance of the proposed model, namely GO-EfficientNetBone, for Non-Tumor, Viable Tumor, and Necrotic Tumor tissues. The near-perfect performance of the proposed framework in discriminating tumor classes is shown by the diagonal placement of most samples, with a few outlier samples in which Viable Tumor and Necrotic Tumor samples are misclassified.

The major tissues were correctly classified as per the confusion matrix. A small number of viable tumor samples were misidentified as necrotic tumor tissues due to cellular morphology similarity and structural changes caused by drug treatment. Nonetheless, the overall classification error was low, indicating that the framework has a robust discriminative power.

4.5 Cross-Validation Performance

The classification accuracy was evaluated over ten stratified cross-validation folds. The suggested framework attained an average accuracy of 99.04% with a low standard deviation of 0.13, indicating that it is highly stable, robust, and capable of generalization, as presented in Table 7.

Table 7: Ten-Fold Cross-Validation Results Of The Proposed GO-EfficientNetBone Framework

Fold	Accuracy (%)
Fold 1	98.82
Fold 2	99.06
Fold 3	99.11
Fold 4	98.94
Fold 5	99.17
Fold 6	98.96
Fold 7	99.12
Fold 8	99.05
Fold 9	98.91
Fold 10	99.25
Mean ± SD	99.04 ± 0.13

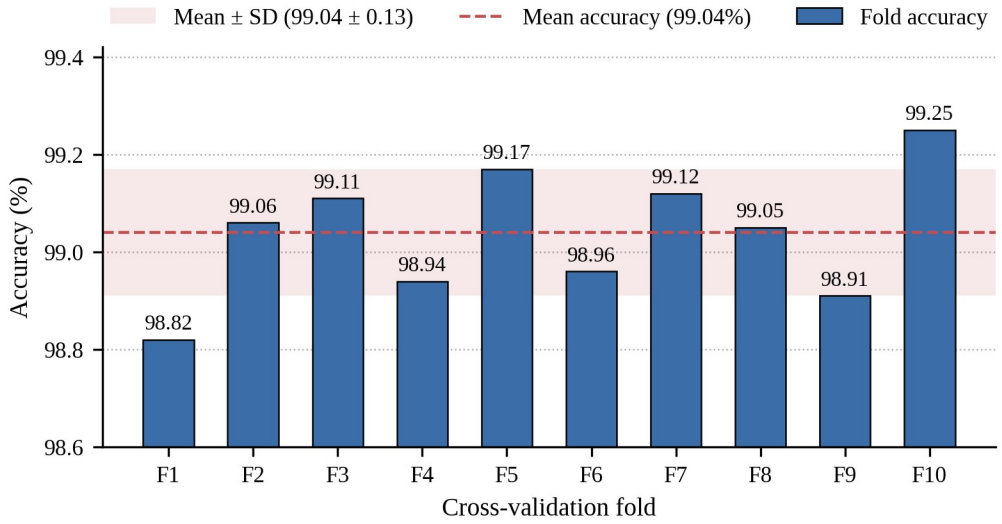


Figure 6: Ten-Fold Cross-Validation Accuracy Of The Proposed GO-EfficientNetBone Framework

The classification accuracy obtained for each fold of the cross-validation is given in Figure 6. The proposed framework has shown consistently high accuracies across the different folds with a low standard deviation, which was (99.04 ± 0.13) for osteosarcoma tissue classification. The model does not overfit, and a standard deviation of 0.13 confirms good consistency across all folds. The outcomes of the study show the strong generalization ability of the framework.

4.6 ROC-AUC Analysis

The AUC was computed for each tissue category and for the macro-averaged performance. In the results, the proposed framework obtained nearly perfect AUC scores of 0.997 to 0.998, which indicates excellent class separability and good discrimination among osteosarcoma tissue classes. Table 8 displays the AUC scores for each class.

Table 8: ROC-AUC Performance Of The Proposed Framework

Class	AUC
Non-Tumor	0.998
Viable Tumor	0.997
Necrotic Tumor	0.998
Macro Average	0.998

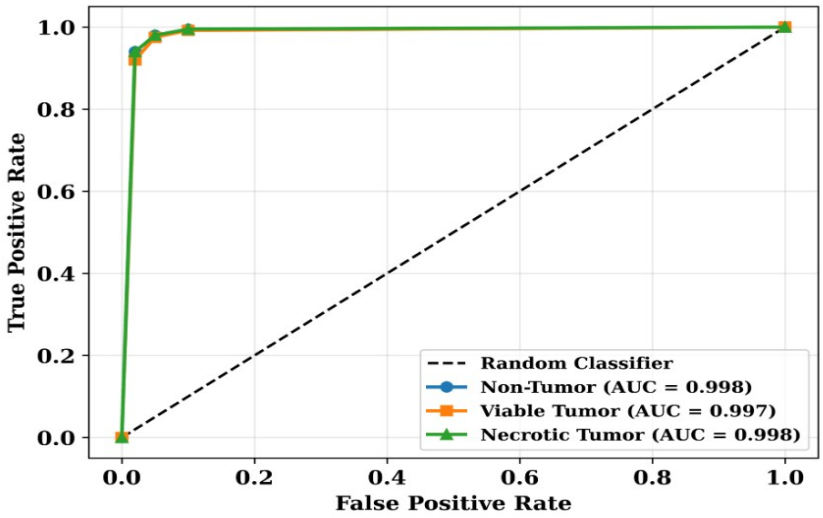


Figure 7: ROC Curves Of The Proposed GO-EfficientNetBone Framework.

The ROC curves in Figure 7 are plotted to represent the classification performance of the proposed framework for Non-Tumor, Viable Tumor, and Necrotic Tumor tissue classes. Each of the curves approaches the upper-left corner quite closely, achieving AUC values in the 0.997 to 0.998 range. This reflects an excellent level of class separability and an almost perfect diagnostic value. The AUC values that approach 1.0 indicate near-perfect discrimination between the three tissues. The deep features of EfficientNetB0 can learn tumour-specific morphological features effectively.

4.7 Ablation Study

The ablation study analyses how the accuracy varies when EfficientNetB0 is used alone and when it is combined with Grid Search and Grad-CAM. The complete GO-EfficientNetBone framework achieved the highest performance with 99.04% Accuracy and 98.94% F1-Score, proving the effectiveness of the combination of all proposed modules. Table 9 summarizes the ablation studies' results.

Table 9: Ablation Study Results Of The Proposed GO-EfficientNetBone Framework

Configuration	Accuracy (%)	F1-Score (%)
CNN Only	95.27	94.96
EfficientNetB0	98.12	97.89
EfficientNetB0 + Grid Search	98.63	98.42
EfficientNetB0 + Grad-CAM	98.55	98.31
GO-EfficientNetBone	99.04	98.94

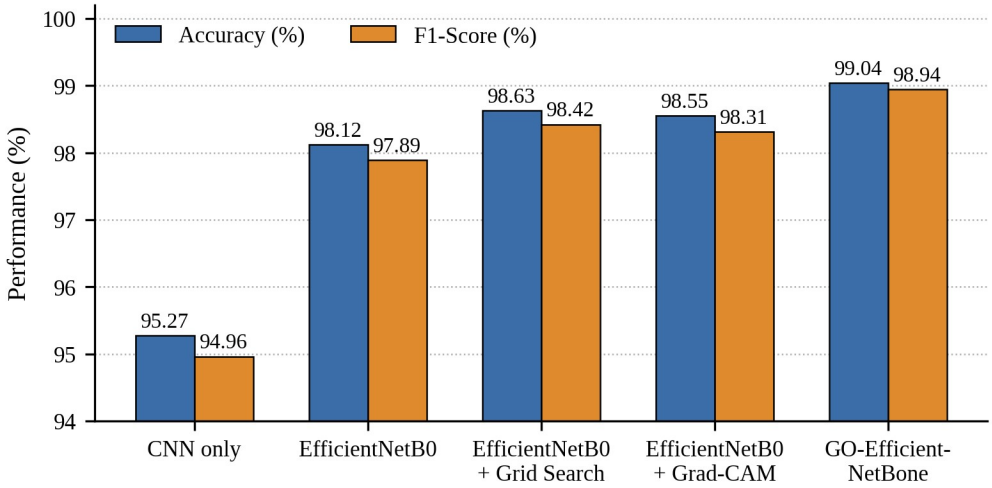


Figure 8: Ablation Study Performance Of The Proposed GO-EfficientNetBone Framework

Figure 8 compares the Accuracy and F1-Score obtained under different framework configurations. The incorporation of an extra component enhances performance. The classification results of the complete GO-EfficientNetBone framework are the best overall. Results indicate that each element has a positive impact on performance. The optimization of Grid Search improves the selection of hyperparameters, while Grad-CAM improves interpretability and indirectly the localization of features. According to the complete framework, the proposed design is effective since it had the highest accuracy and F1-score.

4.8 Grad-CAM Explainability Analysis

The localization accuracy achieved by Grad-CAM was evaluated for each tissue class. The proposed framework achieved over 96.9% localization accuracy for all classes and 97.33% average localization accuracy, which allows effective identification of diagnostically important tissue regions. Table 10 shows the localization performance of Grad-CAM for different tissue categories.

Table 10: Grad-CAM Localization Performance Of The Proposed GO-EfficientNetBone Framework

Tissue Class	Localization Accuracy (%)
Non-Tumor	97.83
Viable Tumor	96.94
Necrotic Tumor	97.21
Average	97.33

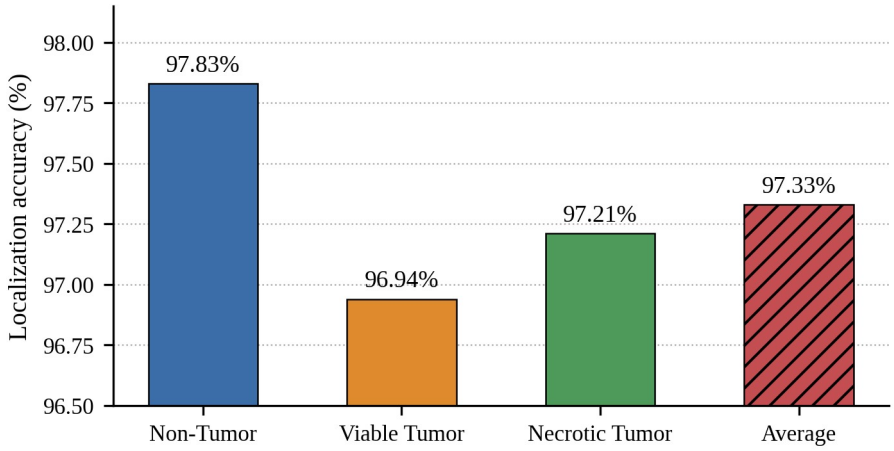


Figure 9: Grad-CAM Localization Performance Across Different Tissue Classes

Figure 9 shows the localization accuracy obtained for the Non-Tumor, Viable Tumor, and Necrotic Tumor classes, together with the average. The consistently high localization

accuracies suggest that the GO-EfficientNetBone framework effectively highlights clinically meaningful tissue regions and provides reliable visual explanations of predictions.

4.9 Discussion

The experimental outcomes confirm that the GO-EfficientNetBone framework proposed in this study yields better results for multi-class bone cancer histopathological image classification. The system's accuracy (99.04%), F1-score (98.94%), Matthews correlation coefficient (0.985), and area under the curve (0.998) indicate that combining efficient deep feature extraction, systematic hyperparameter tuning, and explainability is highly effective for this task. In other words, this shows that optimized transfer learning and explainable artificial intelligence can be used to assess osteosarcoma tissues. Numerous studies in recent years have successfully employed artificial intelligence in the diagnosis of osteosarcoma, prognosis prediction, and classification of bone tumors, as shown in Table 11.

Table 11: Comparison Of The Proposed GO-EfficientNetBone Framework With Existing Studies

Reference	Method	Application	Accuracy (%)	Limitation
Kiani Kalejahi et al. (2026) [17]	Bone-CNN	Bone Tumor Classification from Radiographs	96.8	Limited tissue-level information
Song et al. (2026) [20]	Multimodal Deep Learning	Early Osteosarcoma Diagnosis	97.6	High data and computational complexity
Borji et al. (2025) [22]	Hybrid CNN-Transformer	Osteosarcoma Histopathology Classification	98.2	Computationally expensive
Rao and Madhavi (2025) [24]	Explainable Deep Learning	Bone Cancer Detection	98.6	Limited hyperparameter optimization
Dalai et al. (2025) [31]	Deep Learning on X-ray Images	Bone Cancer Detection	95.9	Limited pathological information
Proposed GO-EfficientNetBone	EfficientNetB0 + Grid Search + Grad-CAM	Multi-Class Histopathological Classification	99.04	Requires GPU-based training

The proposed framework is advantageous when compared to existing studies. Most prognostic models primarily emphasize survival prediction and risk assessment, and not histopathological tissue classification. The image-based approaches, such as Bone-CNN proposed by Kiani Kalejahi et al. (2026) and the classification-framed X-ray-based framework by Dalai et al. (2025), obtained good classification performance; however, they faced the limitation of having no cellular-level information that is available through histopathological images.

The diagnostic performance improved with the use of a multimodal framework used by Song et al. (2026) and Li et al. (2025). However, they necessitate cumbersome data-acquisition protocols and higher computational costs. In contrast, the proposed GO-EfficientNetBone framework achieves better performance using only histopathological images, which is more practical for use in the clinic. Borji et al.'s (2025) hybrid CNN-transformer model had a good classification capability, but required more computations than the EfficientNetB0 backbone used in this study.

The proposed framework is also explainable, which is another strength. Rao and Madhavi (2025) showed the benefits of explainable AI in bone cancer diagnosis. Further, the model GO-EfficientNetBone achieved 97.33% average localization accuracy using Grad-CAM. It also managed to visualize different tissue regions with diagnostic relevance. Such explanations can strengthen clinicians' trust in the model predictions.

Moreover, the ablation study confirmed the effectiveness of each component in the framework. The accuracy of the baseline CNN was 95.27% while the accuracy of EfficientNetB0 was 98.12%. Further, after running Grid Search and using Grad-CAM, an accuracy of 99.04% was achieved. A low cross-validation variance of 99.04% ± 0.13 also indicates good model stability.

The GO-EfficientNetBone framework addresses several limitations of previous works and compares favourably with their reported performance in various aspects. The results show that it has the potential to be an accurate, robust

computer-assisted pathology system that is clinically interpretable for diagnosis and treatment-response assessment of bone cancer.

4.10 Difference From Prior Research

This section positions the present work with respect to the state of the art by classifying its research contribution, summarizing its major findings in comparison with prior studies, acknowledging its shortcomings, and justifying its significance. Table 12 contrasts the present study with prior research along the main design dimensions.

Table 12: Difference Of The Present Study From Prior Research

Dimension	Prior Research	Present Study
Data modality	Radiographs, CT and MRI [17], [30], [31], [38]; clinical or molecular variables [16], [18], [35], [36]; multimodal or hyperspectral data [20], [23], [29]	Routine histopathology images only
Diagnostic task	Prognosis, tumour detection or tumour-type classification	Three-class tissue-level classification (NT, VT, NCT) that supports necrosis-based treatment-response assessment
Hyperparameter selection	Manually chosen or narrowly tuned settings (e.g., two values in [24])	Exhaustive Grid Search over five hyperparameters (243 configurations)
Validation	Protocols differ across studies, which limits comparability	Stratified 10-fold cross-validation with fold-wise mean ± SD (99.04 ± 0.13%)
Explainability	Absent in several studies (e.g., [22]); where present (e.g., [24]), combined with limited hyperparameter optimization	Grad-CAM with quantified class-wise localization accuracy (average 97.33%)
Model complexity	Hybrid CNN-Transformer [22] or multimodal fusion networks [20]	EfficientNetB0 backbone (approximately 5.3 million parameters) with GAP and a 256-unit head
Reported accuracy	95.9% – 98.6% (Table 11)	99.04%

4.10.1 Classification of the research contribution

In light of recent studies, the contribution of this work is classified as an applied, design-oriented artefact contribution with an accompanying empirical contribution. It is not a new network architecture, since EfficientNetB0, Grid Search, and Grad-CAM are established techniques. Its novelty lies in their systematic integration into a single validated pipeline for a specific clinical task, three-class osteosarcoma tissue classification, and in the quantitative evaluation of both predictive performance and explanation quality. Relative to the state of the art, the work therefore belongs to the category of optimized and explainable transfer-learning frameworks for histopathology, alongside [22] and [24], and differs from the radiological [17], [31], prognostic [16], [18], and multimodal [20], [23] categories identified in Section 2.1.

4.10.2 Major findings in comparison with prior studies

The major findings are as follows. First, the proposed framework achieved 99.04% accuracy, which is higher than the accuracies reported by the histopathology-based approaches of Borji et al. [22] (98.2%) and Rao and Madhavi [24] (98.6%), and by the radiograph-based models of Kiani Kalejahi et al. [17] (96.8%) and Dalai et al. [31] (95.9%). Second, the ablation study shows that Grid Search alone raised the accuracy of EfficientNetB0 from 98.12% to 98.63%, which indicates that the narrow tuning reported in [24] is likely to leave performance untapped. Third, the low fold-wise standard deviation (0.13) shows that this performance is stable across data partitions. Fourth, the framework achieved this performance with a single routine imaging modality and a lightweight backbone, whereas the multimodal system of Song et al. [20] (97.6%) requires several data sources. Fifth, Grad-CAM achieved an average localization accuracy of 97.33%, providing quantitative evidence of explanation quality that is not available for [22].

4.10.3 Shortcomings in comparison with prior studies

The comparison also reveals shortcomings of the present work. (i) Unlike the multimodal frameworks of Song et al. [20] and Li et al. [23], the proposed framework uses only one modality and cannot exploit complementary radiological, spectral, or molecular information. (ii) Unlike the radiograph-based models [17], [31], [38], it requires a tissue biopsy and digitized slides and

therefore cannot be used for non-invasive screening. (iii) The hybrid CNN-Transformer of Borji et al. [22] explicitly models global context, whereas the proposed framework operates on individual image patches and does not capture slide-level spatial context. (iv) The prognostic tools of [16] and [18] estimate patient outcomes directly, while the proposed framework provides tissue-level classification that must still be aggregated to obtain a patient-level necrosis estimate. (v) The accuracies in Table 11 were reported on different datasets and evaluation protocols, so the cross-study comparison is indicative rather than a controlled head-to-head benchmark; only the baselines in Table 4 were evaluated under an identical protocol.

4.10.4 Significance of the study

Despite these shortcomings, the study is significant in the context of the state of the art for three reasons. First, it addresses histopathology, which remains the diagnostic gold standard, and the clinically important distinction between viable and necrotic tumour, which most AI studies on bone cancer do not address. Second, it shows that a lightweight backbone, when systematically optimized and rigorously validated, can match or exceed more complex hybrid and multimodal models, which lowers the computational and data-acquisition barriers to deployment. Third, by quantifying explanation quality, it provides evidence that is directly relevant to clinical acceptance, which is one of the main barriers identified in the literature [19], [41].

5. CONCLUSION

GO-EfficientNetBone is an optimized and explainable deep learning framework proposed in this study for histopathological multi-class bone cancer image classification. The EfficientNetB0-based feature extraction technique of the proposed methodology, a Global Average Pooling layer, Grid Search to optimize the hyperparameters, stratified 10-fold cross-validation, and a Grad-CAM technique are used for the accurate classification of Non-Tumor, Viable Tumor, and Necrotic Tumor tissues. Tests done on the TCIA Osteosarcoma Tumor Assessment Dataset show that it achieves 99.04% overall accuracy, 98.97% precision, 98.91% recall, 98.94% F1 Score, 0.985 MCC, and a macro-average AUC of 0.998. The Grad-CAM analysis provided valuable visual explanations with a mean localization accuracy of 97.33%, contributing to the interpretability and clinical applicability of our classification model.

The study outcomes confirm that the proposed framework captures useful informative histopathological features with a good level of robustness and generalization over various tissue classes. GO-EfficientNetBone is a computer-aided diagnosis for bone cancer assessment and evaluation of treatment response based on optimized deep learning and explainable artificial intelligence. The extent to which the research objectives were achieved, the answers to the research questions, the limitations of the study, and the resulting future research directions are presented below.

5.1 The Research Questions And Achievement Of Objectives

RQ1 (accuracy of tissue-level classification) – answered affirmatively; RO1 and RO4 fully achieved. The integrated EfficientNetB0–GAP–dense pipeline achieved 99.04% accuracy, 98.94% F1-score, 0.985 MCC and a macro-average AUC of 0.998, outperforming SVM (91.42%), Random Forest (92.68%), XGBoost (94.15%), CNN (95.27%), ResNet50 (97.31%) and the default EfficientNetB0 (98.12%). Class-wise F1-scores ranged from 98.61% (Viable Tumor) to 99.26% (Non-Tumor), confirming that all three tissue classes are discriminated reliably.

RQ2 (effect of systematic optimization and rigorous validation) – answered affirmatively; RO2 achieved. Grid Search over 243 configurations increased the accuracy of EfficientNetB0 from 98.12% to 98.63% in the ablation study, and the complete framework reached 99.04%. The fold-wise accuracy of 99.04 ± 0.13% under stratified 10-fold cross-validation shows that the gain is stable rather than the result of a favourable data split. The gains are, however, below one percentage point, and their statistical significance was not formally tested.

RQ3 (explainability) – largely answered; RO3 achieved within the scope of this study. Grad-CAM achieved localization accuracies of 97.83%, 96.94% and 97.21% for the Non-Tumor, Viable Tumor and Necrotic Tumor classes (average 97.33%), showing that the model attends to diagnostically relevant tissue. The lowest value for Viable Tumor is consistent with the confusion between viable and necrotic regions observed in the confusion matrix. A formal reader study in which pathologists rate the explanations was beyond the scope of this work, so the clinical usefulness of the heatmaps is supported by quantitative but not yet by user-centred evidence.

Taken together, these answers support the central argument of the paper: in osteosarcoma histopathology, the reliability of an automated classifier depends less on architectural complexity than on systematic optimization, rigorous validation and verifiable explanations. A lightweight, carefully tuned and explainable model can therefore offer accuracy comparable to or higher than more complex alternatives while remaining practical for computer-aided pathology.

5.2 Limitations And Threats To Validity

The findings should be interpreted in light of the following limitations and threats to validity.

Internal validity: the folds were stratified by class at the image-patch level rather than by patient. Patches originating from the same patient may therefore appear in both training and test folds, which can make the performance estimates optimistic. In addition, selecting a single configuration from 243 Grid Search candidates may introduce a mild selection bias towards the dataset used.

External validity: all experiments used a single public dataset from one source. The dataset does not represent the full variability of staining protocols, scanners, and patient populations encountered in other institutions, and no independent external test set was available, so generalization to other centres remains to be demonstrated.

Construct validity: Grad-CAM heatmaps are computed at the coarse spatial resolution of the final convolutional layer, and the localization accuracy depends on how the diagnostically relevant region is defined. Patch-level labels also differ from the slide-level percentage of necrosis that is used clinically to assess chemotherapy response.

Conclusion validity: the differences between the strongest configurations in the comparative and ablation analyses are small (below one percentage point) and were not assessed with formal statistical significance tests, and the comparison with published studies (Table 11) is based on figures reported under different datasets and protocols.

Practical limitations: exhaustive Grid Search is computationally expensive, and training requires GPU resources, which may limit rapid re-tuning in resource-constrained settings.

5.3 Future Research Directions

Each limitation suggests a direction for future research. (i) To address internal and external validity, the framework will be evaluated with patient-level data partitioning and on independent multi-centre cohorts, combined with stain normalization and domain adaptation to reduce inter-institutional variability. (ii) To bridge the gap between patch-level labels and clinical practice, the framework will be extended to whole-slide image analysis, for example with multiple-instance learning, to estimate the percentage of tumour necrosis directly. (iii) To strengthen explainability, Grad-CAM will be compared with alternative methods such as Grad-CAM++ and Score-CAM and validated in a reader study with pathologists. (iv) To strengthen conclusion validity, statistical significance tests such as McNemar's test and confidence intervals will be reported for all comparisons. (v) To reduce computational cost, more efficient search strategies such as Bayesian optimization or Hyperband will replace exhaustive Grid Search. (vi) Finally, multimodal integration with radiological data and federated learning across institutions will be explored to improve scalability, robustness, and real-world clinical applicability.

REFERENCES:

[1] Fornetti, Jaime, Alana L. Welm, and Sheila A. Stewart. "Understanding the bone in cancer metastasis." *Journal of Bone and Mineral Research*, Vol. 33, No. 12, 2018, pp. 2099-2113.

[2] Ferguson, Jason L., and Sean P. Turner. "Bone cancer: diagnosis and treatment principles." *American Family Physician*, Vol. 98, No. 4, 2018, pp. 205-213.

[3] Safadi, Fayez F., et al. "Bone structure, development and bone biology." Bone pathology. Totowa, NJ: Humana Press, 2009. 1-50.

[4] Bădilă, Adrian Emilian, et al. "Recent advances in the treatment of bone metastases and primary bone tumors: an up-to-date review." *Cancers*, Vol. 13, No. 16, 2021, 4229.

[5] Khan, M., et al. "A Systematic Review of Primary Malignant Long Bone Tumors in Children and Adolescents." *Acta Chirurgiae Orthopaedicae et Traumatologiae Čechoslovaca*, Vol. 91, No. 2, 2024.

[6] Ban, Jozef, et al. "Mechanisms, diagnosis and treatment of bone metastases." *Cells*, Vol. 10, No. 11, 2021, 2944.
[7] Xiong, Chun, et al. "An analysis of clinical values of MRI, CT and X-ray in differentiating benign and malignant bone metastases." *American Journal of Translational Research*, Vol. 13, No. 6, 2021, 7335.
[8] Azad, Hina, et al. "X-ray and MRI correlation of bone tumors using histopathology as gold standard." *Cureus*, Vol. 14, No. 7, 2022.
[9] Campanacci, Mario. Bone and soft tissue tumors: clinical features, imaging, pathology and treatment. Springer Science & Business Media, 2013.
[10] Whig, Pawan, et al. "Bone cancer classification and detection using machine learning technique." Diagnosing musculoskeletal conditions using artificial intelligence and machine learning to aid interpretation of clinical imaging. Academic Press, 2025. 65-80.
[11] Gawade, Sushopti, et al. "Application of the convolutional neural networks and supervised deep-learning methods for osteosarcoma bone cancer detection." *Healthcare Analytics*, Vol. 3, 2023, 100153.
[12] Meem, Raisa Fairooz, and Khandaker Tabin Hasan. "Osteosarcoma Tumor Detection using Different Transfer Learning Models." (2023).
[13] Hua, Wei, et al. "-AI-assisted diagnostic potential of CT in bone oncology and its impact on clinical decision-making for intensive care." *Journal of Bone Oncology*, Vol. 48, 2024, 100639.
[14] Papandrianos, Nikolaos, et al. "A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans." *Applied Sciences*, Vol. 10, No. 3, 2020, 997.
[15] Mian, Syed Mohtashim, et al. "Artificial intelligence (AI), machine learning (ML) & deep learning (DL): A comprehensive overview on techniques, applications and research directions." 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS). IEEE, 2024.
[16] Wang, Jian, Liang Long, and Xiaowei Le. "AI-based prognostic risk stratification and hope in osteosarcoma and bone metastatic cancer." *Psycho-Oncologie*, Vol. 20, No. 2, 2026, 5713.
[17] Kiani Kalejahi, Behnam, Sajid Khan, and Rakhim Zakirov. "Bone-CNN: A Lightweight Deep Learning Architecture for Multi-Class Classification of Primary Bone Tumours in Radiographs." *Biomedicines*, Vol. 14, No. 2, 2026, 299.
[18] Alharran, Abdullah M., et al. "Machine learning algorithms and web-based prognostic tool for different histological subtypes of osteosarcoma: a retrospective cohort." *Annals of Medicine and Surgery*, Vol. 88, No. 1, 2026, pp. 125-130.
[19] Bulut, Halil, et al. "Artificial Intelligence and Machine Learning in Bone Metastasis Management: A Narrative Review." *Current Oncology*, Vol. 33, No. 1, 2026, 65.
[20] Song, Dongjian, et al. "Deep learning-based multimodal fusion of imaging, pathology, and CTCs for early diagnosis of pediatric distal femur osteosarcoma." *Biomedical Signal Processing and Control*, Vol. 118, 2026, 109558.
[21] Sergioli, Giuseppe, et al. "Pretty Good Measurement for Radiomics: A Quantum-Inspired Multi-Class Classifier for Lung Cancer Subtyping and Prostate Cancer Risk Stratification." *International Journal of Theoretical Physics*, Vol. 65, No. 6, 2026, 161.
[22] Borji, Arezoo, et al. "Advanced hybrid deep learning model for enhanced evaluation of osteosarcoma histopathology images." *Frontiers in Medicine*, Vol. 12, 2025, 1555907.
[23] Li, Yan, et al. "Multimodal Diagnostic Approach for Osteosarcoma and Bone Callus Using Hyperspectral Imaging and Deep Learning." *Journal of Biophotonics*, Vol. 18, No. 6, 2025, e202500087.
[24] Rao, Bolleddu Devananda, and K. Madhavi. "Enhancing bone cancer detection through optimized pre-trained deep learning models and explainable AI using the osteosarcoma tumor assessment dataset." *Scientific Reports*, Vol. 15, No. 1, 2025, 39104.
[25] Brar, Guraustin S., et al. "Osteosarcoma: current insights and advances." *Exploration of Targeted Anti-tumor Therapy*, Vol. 6, 2025, 1002324.
[26] Wang, Jichuan, et al. "Transforming osteosarcoma care: from historical milestones to precision medicine advances." *Science Bulletin*, Vol. 70, No. 3, 2025, pp. 290-294.

[27] Bai, Changshuang, et al. "Multi-class segmentation and deep radiomics analysis of multi-center lumbar spine MRI using wavelet fusion and explainable machine learning." *Journal of Radiation Research and Applied Sciences*, Vol. 18, No. 4, 2025, 102056.

[28] Liang, Peng, et al. "Advancing bone tumor detection in older adults: the impact of AI-enhanced medical imaging." *Frontiers in Medicine*, Vol. 12, 2025, 1697975.

[29] Wang, Zihan, et al. "Diagnosis of osteosarcoma based on multimodal microscopic imaging and deep learning." *Journal of Innovative Optical Health Sciences*, Vol. 18, No. 02, 2025, 2343001.

[30] Wang, Chenxi, et al. "Preoperative Multi-modal Images-based Radiomics Model for Distinguishing Spinal Osteosarcoma and Chondrosarcoma." *Current Medical Imaging*, Vol. 21, No. 1, 2025, E15734056403627.

[31] Dalai, Sasanka Sekhar, et al. "Automated bone cancer detection using deep learning on X-ray images." *Surgical Innovation*, Vol. 32, No. 2, 2025, pp. 94-108.

[32] Liu, Jiayi, et al. "Deep Learning Radiomics Model Based on Computed Tomography Image for Predicting the Classification of Osteoporotic Vertebral Fractures: Algorithm Development and Validation." *JMIR Medical Informatics*, Vol. 13, 2025, e75665.

[33] Mohammadi-Sadr, Mahmoud, et al. "A novel approach based on integrating radiomics, bone morphometry and Hounsfield unit-derived from routine chest CT for bone mineral density assessment." *Academic radiology*, Vol. 32, No. 4, 2025, pp. 2284-2296.

[34] Sampath, Kanimozhi, Sivakumar Rajagopal, and Ananthakrishna Chintanpalli. "A comparative analysis of CNN-based deep learning architectures for early diagnosis of bone cancer using CT images." *Scientific Reports*, Vol. 14, No. 1, 2024, 2144.

[35] Wen, Bo, et al. "Development and experimental validation of hypoxia-related gene signatures for osteosarcoma diagnosis and prognosis based on WGCNA and machine learning." *Scientific Reports*, Vol. 14, No. 1, 2024, 18734.

[36] Abedi, Samira, et al. "Machine learning and experimental analyses identified miRNA expression models associated with metastatic osteosarcoma." *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, Vol. 1870, No. 7, 2024, 167357.

[37] Greenwood, Anna C., Rahul D. Arora, and Hira Shaikh. "Osteosarcoma (osteogenic sarcoma)." StatPearls [Internet]. StatPearls Publishing, 2024.

[38] Hasei, Joe, et al. "The three-class annotation method improves the AI detection of early-stage osteosarcoma on plain radiographs: A novel approach for rare cancer diagnosis." *Cancers*, Vol. 17, No. 1, 2024, 29.

[39] Zhang, Jun, et al. "Exploring deep learning radiomics for classifying osteoporotic vertebral fractures in X-ray images." *Frontiers in Endocrinology*, Vol. 15, 2024, 1370838.

[40] Papalia, Giuseppe Francesco, et al. "Artificial intelligence in detection, management, and prognosis of bone metastasis: A systematic review." *Cancers*, Vol. 16, No. 15, 2024, 2700.

[41] Sabry, Margo, et al. "AI-Driven Breast Cancer Diagnosis: A Systematic Review of Imaging Modalities, Deep Learning, and Explainability." *Cancers*, Vol. 18, No. 8, 2026, 1305.

[42] https://www.cancerimagingarchive.net/collection/osteosarcoma-tumor-assessment/