

DNA-Inspired Symbolic Encoding with Adaptive Block-level Hybrid CGR-FCGR Features for Osteoporosis Progression Detection

Lipika Dinda

National Institute of Technology Jamshedpur

Jitesh Pradhan

National Institute of Technology Jamshedpur

Manish Raj

All India Institute of Medical Science

Tapan Kumar Dey

`tapan.dey@jaipur.manipal.edu`

Manipal University Jaipur

Research Article

Keywords: Osteoporosis Classification, CGR-FCGR Feature Fusion, Symbolic Encoding, Transformer-Based Feature Learning, HTGCA-Net

Posted Date: July 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9971525/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

DNA-Inspired Symbolic Encoding with Adaptive Block-level Hybrid CGR-FCGR Features for Osteoporosis Progression Detection

Lipika Dinda¹, Jitesh Pradhan², Manish Raj³,
Tapan Kumar Dey^{4*}

^{1,2}CVPR Lab, Department of Computer Science and Engineering, NIT Jamshedpur, Adityapur, 831014, Jharkhand, India.

³Department of Orthopaedics, All India Institute of Medical Sciences, Deoghar, Rampur, 814152, Jharkhand, India.

^{4*}Department of Computer Science and Engineering, Manipal University Jaipur, Jaipur, 303007, Rajasthan, India.

*Corresponding author(s). E-mail(s): tapan.dey@jaipur.manipal.edu;

Contributing authors: lipikadindanitjsr@gmail.com;

jiteshpradhan.cse@nitjsr.ac.in;

manish.orthopaedics@aiimsdeoghar.edu.in;

Abstract

Osteoporosis is a progressive skeletal disorder characterized by reduced bone mineral density (BMD) and deterioration of trabecular microarchitecture, increasing fracture risk. Although Dual-Energy X-ray Absorptiometry (DXA) is the clinical gold standard, its limited accessibility motivates the development of cost-effective screening approaches using plain radiographs. In this research, a novel framework is proposed that integrates entropy-guided adaptive block division with DNA-inspired symbolic encoding and Chaos Game Representation (CGR) and Frequency CGR (FCGR) to capture complementary spatial and frequency-based characteristics of bone texture. The entropy-guided block division enables region-specific analysis by partitioning the image into structurally informative segments, thereby enhancing the representation of local trabecular variations while reducing the influence of homogeneous or less relevant regions. The resulting hybrid feature representation is refined through a Transformer-based projection mechanism to obtain compact and discriminative embeddings. For Binary class and Multiclass classification task, a hybrid deep learning model HTGCA-Net (Hybrid

Transformer GRU Context Attention Network), is developed by integrating convolutional layers, Transformer encoders, Bidirectional GRU units, and attention with context. The method is evaluated on a clinically validated dataset collected under an ICMR project, with labels derived from DXA-based WHO T-score criteria. Experimental results demonstrate that the proposed approach achieves an accuracy of 0.98 for binary class classification and 0.92 for multiclass classification under an 80:20 patient-wise split, while maintaining stable performance across different data distributions. These findings highlight its potential for reliable and accessible osteoporosis screening

Keywords: Osteoporosis Classification, CGR-FCGR Feature Fusion, Symbolic Encoding, Transformer-Based Feature Learning, HTGCA-Net

1 Introduction

Osteoporosis is a progressive skeletal disorder characterized by reduced bone mineral density (BMD) and deterioration of bone microarchitecture, leading to increased fracture risk [1], [2]. It is considered a major public health concern, particularly among aging populations, where fragility fractures significantly impact morbidity, mortality, and healthcare costs [2]. Therefore, early diagnosis is crucial for successful clinical therapy. The standard method for assessing BMD is dual-energy X-ray absorptiometry (DXA [3]; however, its restricted availability, high cost, and need for specialized infrastructure limit its widespread usage, particularly in low-resource settings. As a result, there is an increasing demand for alternative, affordable screening methods that can make use of regularly obtained image modalities.

X-rays, or plain radiographs, are easily accessible and provide important structural details on trabecular bone structures. There is diversity in diagnosis due to the subjectivity and reliance on clinical skill of traditional evaluation methods such as visual grading techniques [4]. Machine learning and deep learning techniques have been investigated more and more for automated osteoporosis detection [5] in order to overcome these constraints. Although transformer-based models and convolutional neural networks have demonstrated promise performance [6, 7], these methods frequently lack interpretability and require huge annotated datasets. Furthermore, many current techniques concentrate mostly on global representations, thereby ignoring minute local structural differences that are essential for early diagnosis.

From a structural standpoint, complex alterations in trabecular bone patterns with fractal-like features are linked to osteoporosis [8, 9]. Radiographic textures reflect these microstructural changes, which can be successfully represented with the right feature representations. Both local abnormalities and global structural distributions may not be sufficiently represented by conventional texture descriptors or global statistical measurements. Therefore, to increase classification efficiency and robustness, a more complete feature representation that incorporates several complementing features is needed.

In this research, a hybrid feature extraction framework is proposed that combines Chaos Game Representation (CGR) and Frequency Chaos Game Representation

(FCGR) for osteoporosis classification from hip X-ray images. The CGR-based features capture fine-grained spatial and fractal patterns by encoding local intensity transitions into a structured representation [10], while FCGR-based features quantify the frequency distribution of these patterns, thereby providing a global statistical perspective [11]. The suggested hybrid technique seeks to produce a more robust and discriminative representation of bone texture by combining these two complementing feature sets.

Table 1: List of abbreviations

Abbreviation	Full form
BMD	Bone Mineral Density
DEXA / DXA	Dual-Energy X-ray Absorptiometry
WHO	World Health Organization
DCNN	Deep Convolutional Neural Network
ANN	Artificial Neural Network
DL	Deep Learning
ML	Machine Learning
SVM	Support Vector Machine
DT	Decision Tree
KNN	k-Nearest Neighbours
RF	Random Forest
AUC	Area Under the Curve
LDA	Linear discriminant analysis
MCI	Mandibular Cortex Index
LDCT	low-dose computed tomography
RTA	Radiomic Texture Analysis
CGR	Chaos Game Representation
FCGR	Frequency Chaos Game Representation
HTGCA-Net	Hybrid Transformer GRU Context Attention Network
A_c	Accuracy
S_n	Sensitivity
S_p	Specificity
P_r	Precision
R_c	Recall
ROC	Receiver Operating Characteristic
F_s	F1-score
Avg	Average

2 Literature Survey

Osteoporosis is a progressive skeletal disorder characterized by reduced bone mineral density (BMD) and deterioration of bone microarchitecture, leading to an increased risk of fractures. Despite being the clinical gold standard for diagnosis, Dual-Energy X-ray Absorptiometry (DXA) is not widely used due to its high cost, restricted accessibility, and incapacity to provide precise structural bone information. As a result, current research has concentrated more on creating artificial intelligence (AI)-based

techniques for early and economical osteoporosis detection using clinical data and medical imaging.

Several researches have explored machine learning techniques[19] [24] based on clinical variables and BMD measurements for early risk prediction[23]. For instance, hybrid expert systems combining multiple classifiers such as Logistic Regression, Decision Trees, Random Forests, and XGBoost have been proposed to improve predictive performance[24]. Similarly, comparative analyses involving multiple machine learning models trained on DXA-derived features and clinical indicators have demonstrated improved diagnostic capability over traditional screening tools[33]. However, these approaches are primarily dependent on structured data and often fail to capture spatial and structural characteristics inherent in medical images.

To address these limitations, deep learning approaches utilizing advanced imaging modalities such as CT and MRI scan6 have been widely investigated. Convolutional neural networks (CNNs) have been employed for vertebral segmentation, BMD estimation, and fracture detection, achieving high accuracy and strong correlation with clinical measurements[20],[22],[26]. Ensemble CNN models and transfer learning-based architectures have further improved performance, particularly in detecting vertebral abnormalities[25],[27],[28]. Additionally, radiomics-based methods integrating feature extraction from MRI with machine learning classifiers have shown promising results for multiclass classification of bone conditions[29]. Despite their effectiveness, these methods are constrained by high imaging costs, limited availability, and, in the case of CT scan, increased radiation exposure.

In recent years, X-ray imaging has emerged as a practical alternative due to its low cost, wide availability, and minimal radiation dose[32],[41],[43]. Several researches have demonstrated the effectiveness of deep learning models, including EfficientNet, Inception-based architectures, and ensemble CNN frameworks, for osteoporosis detection using radiographs[42],[25]. Multi-task learning models and segmentation-based approaches[29],[31] have also been proposed to simultaneously detect fractures and classify bone conditions. Furthermore, classical image processing techniques combined with machine learning classifiers have shown competitive performance in certain scenarios, particularly when computational resources are limited.

Beyond binary classification, multiclass classification involving normal, osteopenia, and osteoporosis categories has gained increasing attention due to its clinical relevance. Advanced architectures incorporating multi-scale feature fusion, channel attention mechanisms, and joint learning frameworks have been developed to improve classification sensitivity and accuracy. Recent researches have also explored automated localization and staging methods using object detection models, enabling more efficient and scalable diagnostic systems.

Zhang *et al.*[41] introduced a unified learning framework for osteoporosis analysis that integrates localization, segmentation, and classification within a single pipeline. Their approach employs boundary-aware heatmaps and gated convolution mechanisms to enhance feature fusion and contextual representation for improved vertebral assessment. Rasool *et al.*[25] proposed KONet, a weighted ensemble architecture that integrates multiple convolutional neural networks to enhance the detection of subtle

structural variations in knee X-ray images for osteoporosis diagnosis. He *et al.*[29] utilized radiomic feature extraction from segmented T1-weighted and T2-weighted MRI images of the lumbar spine, followed by feature selection and classification models to distinguish among normal, osteopenia, and osteoporosis conditions. Peng *et al.*[38] employed VB-Net for trabecular bone segmentation and DenseNet for classification and regression tasks, enabling both osteoporosis detection and bone density prediction from CT images. Similarly, Chen *et al.*[36] combined vertebral segmentation with radiomic texture analysis to construct a hierarchical classification framework capable of identifying abnormal bone density and further differentiating between osteopenia and osteoporosis. Ozic *et al.*[37] proposed a YOLOv5-based system for automated localization and staging of osteoporosis from panoramic radiographs. Recent researches in related imaging domains have further highlighted the effectiveness of attention-driven architectures. For instance, Li *et al.*[23] proposed a multiview attention network with adaptive augmentation strategies to improve generalization under limited data conditions.

Two summary tables are provided to offer an organized comparison of current methods. While Table 2 describes representative research concentrated on binary class classification (Normal vs. Osteoporosis), Table 3 covers approaches for multiclass classification (Normal, Osteopenia, and Osteoporosis). In addition to highlighting the variety of imaging modalities, feature extraction methods, and model architectures used in recent researches, these tables also show common issues with data dependency, generalization, and interpretability.

The current researches focus on using widely accessible X-ray imaging as an affordable substitute for osteoporosis testing in order to overcome these restrictions. Specifically, by combining Chaos Game Representation (CGR) and Frequency Chaos Game Representation (FCGR), a hybrid feature extraction approach is presented that allows for the modeling of both spatial structural patterns and frequency-based statistical properties of trabecular bone. When compared to traditional handcrafted or deep features alone, these complementing representations offer a more thorough description of bone texture. Additionally, a transformer-guided deep learning architecture that incorporates convolutional layers, sequential modeling components, and attention mechanisms is used to handle the extracted hybrid features. This design maintains a fair trade-off between interpretability and model complexity while enabling efficient learning of both local and global dependencies. Consequently, the suggested approach provides a reliable and scalable osteoporosis classification solution with enhanced generalization across different data distributions.

3 Motivation and Contribution

Early detection of osteoporosis from plain radiographic images remains a challenging task due to the subtle structural variations in trabecular bone patterns and the limited visibility of density changes in conventional X-ray imaging. Although several feature extraction and deep learning approaches have been proposed, many existing methods either rely on fixed spatial partitioning strategies or directly process pixel-level information without effectively capturing symbolic or structural relationships

Table 2: Summary of Osteoporosis Diagnosis methods for Binary Class(Healthy and Osteoporosis)

Author, year	Modality	Classification Type	Methods	Performance	Key Findings
Valentinitsch <i>et al.</i> [17], 2019	CT scan(Lumbar Spine)	Binary Class	3D Texture Features (HAR, HoG, LBP, Wavelets) + Classification (Random Forest)	AUC: 0.88	An automated CT-based pipeline using 3D texture features and regional volumetric bone mineral density(vBMD) improves discrimination of vertebral fracture risk, outperforming vBMD alone.
Bhattacharya <i>et al.</i> [18], 2019	Bone X-ray	Binary Class	SVM, NN	A _c : 0.95	A cost-effective automated method accurately classifies osteoporosis, showing potential for real-time detection in clinical settings.
Shim <i>et al.</i> [19], 2020	Clinical Data	Binary Class	KNN, DT, RF, GBM, SVM, ANN, LR with/without feature selection	ANN AUROC: 0.743	ANN outperformed others; good for early screening
Mu <i>et al.</i> [20], 2021	CT scan, DXA scan	Binary Class	Deep CNN using body fat ratio and BMD; validated via DXA	A _c : 0.89	Body fat ratio had strong BMD correlation; CT-scan consistent with DXA
Yabu <i>et al.</i> [21], 2021	MRI	Binary Class	Ensemble of 9 CNNs (VGG16, VGG19, DenseNet201, ResNet50, etc.)	AUC: 0.949, S _n : 0.88, S _p : 0.87	CNNs match spine surgeons in fresh OVf detection
Dzierzak <i>et al.</i> [22], 2022	CT scan (L1 vertebra)	Binary Class	6 DCNNs including VGG16, ResNet50 (transfer learning)	VGG16: A _c : 0.95, TPR: 0.96, TNR: 0.94	Transfer learning effective on small CT dataset
Liu <i>et al.</i> [23], 2022	CT scan(Lumbar Spine) + Clinical Data	Binary Class	ML models (LR, SVM, ANN, RF, Radial-basis function kernel)	ROC of hierarchical LR= 0.818	A CT + clinical databased ML model enables opportunistic osteoporosis screening without additional cost, allowing early risk identification even without DXA availability.
Chen <i>et al.</i> [36], 2023	Chest LDCT	Binary Class	ResNet50 for segmentation + hierarchical SVM classifier	AUC: 0.96, A _c : 0.91	ML + RTA accurately predicts bone density from LDCT, enabling opportunistic screening.
Albuquerque <i>et al.</i> [24], 2023	DXA scan	Binary Class	ML models (RF, XGBoost, LR, etc.)	S _n of RF : 0.853, S _p of RF : 0.879, F _s of RF : 0.859	Random Forest model combining Osseus measurements with clinical factors (age, BMI) effectively screens osteoporosis with high sensitivity and specificity, offering a low-cost alternative to DXA for early detection.
Suh <i>et al.</i> [26], 2023	Femur of Neck X-ray + Clinical Data	Binary Class	Customized DL model	A _c : 0.92	An interpretable deep learning model using clinical features outperforms conventional methods and, through explainable AI, provides clear feature-level explanations for individualized osteoporosis risk assessment.
Rasool <i>et al.</i> [25], 2024	Knee X-ray	Binary Class	Ensemble CNN (EfficientNetB0 + DenseNet121)	A _c : 0.96, F1: 0.97	Robust detection of knee Osteoporosis via ensemble CNN
Belali <i>et al.</i> [27], 2025	DXA scan	Binary Class	Custom DCNN	A _c : 0.98; P _r : 0.98; R _c : 0.98; F _s : 0.98	CNN-based model enables fast and accurate fracture detection from X-ray images in osteoporotic patients, outperforming traditional architectures and supporting efficient, low-cost diagnosis.
Srinivasan <i>et al.</i> [28], 2025	Knee X-ray	Binary Class	Cascaded ShuffleNet + ANN	A _c : 0.97	BoneVoyage combines ShuffleNet and ANN to enable fast, accurate, and reliable early osteoporosis detection from bone density images.
Dinda <i>et al.</i> [43], 2025	Knee X-ray	Binary Class	Feature Extraction + DL based hybrid classification(ResNET, MLP, Autoencoder)	A _c : 0.90, P _r : 0.88, R _c :0.84 , F _s :0.86	Hybrid modeling combines handcrafted (local structural) and ViT (global contextual) features in osteoporosis classification.

within bone texture patterns. To address these limitations, this research introduces a biologically inspired representation framework that combines entropy-guided spatial analysis, DNA-based symbolic encoding, Chaos Game Representation(CGR), Frequency Chaos Game Representation(FCGR), and a hybrid deep learning classifier for robust osteoporosis classification.

The primary contributions of this research are summarized as follows:

Table 3: Summary of Multiclass(Healthy, Osteopenia and Osteoporosis) Osteoporosis Diagnosis Methods by researchers in various years

Author, year	Modality	Classification Type	Methods	Performance	Key Findings
He <i>et al.</i> [29], 2020	Lumbar Spine MRI	Multiclass	Segmentation + Classification	Osteopenia (A_c : 0.74, S_n : 0.68, S_p : 0.78, AUC: 0.81), Osteoporosis (A_c : 0.78, S_n : 0.87, S_p : 0.68, AUC: 0.79)	MRI-based radiomic models can effectively distinguish BMD categories; useful for osteoporosis detection.
Zhang <i>et al.</i> [41], 2020	Lumbar Spine X-ray	Multiclass	Deep CNN based classification	AUC(Osteoporosis) : 0.76, AUC(Osteopenia) : 0.81	A DCNN can classify osteoporosis and osteopenia from lumbar X-rays, indicating potential for low-cost screening without DXA.
Prakash <i>et al.</i> [30], 2021	T-scores from bone density test	Multiclass	4x Expert System ML framework	A_c : 0.75-0.97	Multi-model system showed high predictive range
Fang <i>et al.</i> [31], 2021	CT scan(Spinal and Abdominal)	Multiclass	Segmentation (U-Net) + BMD calculation (DenseNet-121)	BMD1 : 0.992, BMD2 : 0.986, BMD3 : 0.980	Deep learning accurately segments vertebrae and estimates BMD, enabling effective opportunistic osteoporosis screening from CT scans.
Dong <i>et al.</i> [32], 2022	Spine X-ray	Multiclass	GoogLeNet	AUC-ROC: 0.99, S_n : 0.59, PPV: 0.91, F_s : 0.72	GoogLeNet accurately classifies vertebral fractures. High AUC indicates potential for automated OCF detection.
Fasihi <i>et al.</i> [33], 2022	DEXA Scan	Multiclass	8 ML models (DT, RF, KNN, SVM, GB, Extra Trees, AdaBoost, MLP)	AUROC: GB (0.95) in women, RF (0.91) in men; RF best for exercise (0.96 women, 0.99 men)	AI models show high classification accuracy. RF best for treatment guidance. Gender-specific modeling improves clinical relevance.
Xue <i>et al.</i> [34], 2022	CT scan(Lumbar Spine)	Multiclass	DSNet with channel attention, dual-scale fusion, quadratic weight computation	A_c : 0.83, R_c : 0.90	DSNet improves classification accuracy and sensitivity; channel attention enhances early Osteoporosis detection.
Zhang <i>et al.</i> [35], 2023	X-ray	Multiclass	Segmentation with DarkNet-53 + ResNet-101 classifier	A_c : 0.93, AUC: Normal: 0.973, Osteopenia: 0.965, Osteoporosis: 0.985	Feature pyramid model reduces annotation cost; excellent diagnostic results across three classes.
Chen <i>et al.</i> [36], 2023	Chest LDCT	Multiclass	ResNet50 for segmentation + hierarchical SVM classifier	AUC: 0.98, A_c : 0.94	ML + RTA accurately predicts bone density from LDCT, enabling opportunistic screening.
Ozic <i>et al.</i> [37], 2023	Dental Panoramic Radiograph	Multiclass	YOLOv5 for automatic localization and staging via MCI index	Avg P_r : 0.994, R_c : 0.993, F_s : 0.993, Inference time: 14.3 ms	High performance in staging OP using panoramic X-rays; no ROI cropping needed.
Peng <i>et al.</i> [38], 2024	CT scan	Multiclass	VB-Net for trabecular segmentation + DenseNet for classification and regression	AUC: 0.933, F_s : 0.798, R^2 : 0.878, r : 0.94, MAE: 10.26, RMSE: 13.51	Strong generalization across datasets; models predict BMD and classify Osteoporosis accurately, reducing radiation/-cost.
Pan <i>et al.</i> [42], 2024	CT scan(Lumbar Spine) + Clinical Data	Multiclass	Multi-feature fusion DCNN	AUC : 0.98	The multi-feature DCNN model using L1 vertebra and clinical data enables three-class bone density classification and BMD estimation, supporting large-scale opportunistic osteoporosis screening.
Sarmadi <i>et al.</i> [39], 2024	Knee X-ray	Multiclass	Vision Transformer (ViT) vs CNNs (VGG16, ResNet50)	ViT A_c : 0.61	ViT outperformed CNNs with large data
Wu <i>et al.</i> [40], 2025	Low-Dose Chest CT Scan	Multiclass	Deep learning for vertebral segmentation + BMD estimation vs QCT reference	$S_n > 0.86$, $S_p > 0.98$	Accurately measures BMD from T12-L2 on LDCT; generalizes well across scanners; useful for mass screening.
Dinda <i>et al.</i> [43], 2025	Knee X-ray	Multiclass	Feature Extraction + DL based hybrid classification(ResNET, MLP, Autoencoder)	A_c : 0.90, P_r : 0.89, R_c :0.90 , F_s : 0.90	Hybrid modeling combines handcrafted (local structural) and ViT (global contextual) features in osteoporosis classification.

1. **Entropy-guided adaptive spatial partitioning for trabecular structure analysis:** An adaptive block division strategy based on Shannon entropy is introduced to dynamically identify structurally complex regions within hip X-ray images. Unlike conventional uniform or grid-based partitioning methods, the proposed approach selectively refines regions exhibiting higher textural irregularity, enabling the model to focus on diagnostically relevant bone structures such as trabecular thinning, cortical discontinuity, and localized density variations.

2. **DNA-inspired symbolic encoding of grayscale intensity patterns:** A DNA-inspired symbolic encoding strategy is proposed to convert grayscale pixel intensities into nucleotide-like symbolic sequences, enabling sequence-based fractal representations such as CGR and FCGR for bone texture analysis.
3. **CGR and FCGR-based multi-domain feature representation:** The symbolic sequences generated from the encoded X-ray images are analyzed using Chaos Game Representation (CGR) and Frequency Chaos Game Representation (FCGR). While CGR captures spatial recurrence and fractal characteristics of intensity transitions, FCGR quantifies statistical frequency distributions of symbolic sub-sequences. The complementary nature of these two representations provides a multi-level description of trabecular bone morphology.
4. **Fusion of spatial and frequency symbolic features for enhanced discriminative capability:** Features derived from CGR and FCGR representations are combined to form a comprehensive descriptor that integrates spatial recurrence patterns with frequency-based symbolic statistics. This fusion strategy enables the proposed framework to capture subtle morphological differences between normal, osteopenic, and osteoporotic bone structures.
5. **A hybrid deep learning architecture for robust osteoporosis classification:** To effectively learn complex relationships among the extracted features, A hybrid deep learning classifier integrating Conv1D, Transformer encoder, BiGRU, and context-attention is developed to learn hierarchical dependencies within the extracted CGR and FCGR features. This architecture allows the model to simultaneously capture local feature interactions, global dependencies, and discriminative feature importance, leading to improved performance in both binary and multiclass osteoporosis classification tasks.

Through the integration of entropy-driven spatial analysis, symbolic sequence modeling, fractal-based representations, and hybrid deep learning classification, the proposed framework provides a comprehensive methodology for capturing subtle bone structural variations in radiographic images, thereby improving the reliability of osteoporosis detection.

4 Proposed Methodology

The overall framework is organized into three stages, as illustrated in Figure 1. First, the dataset is described in Section 4.1, ensuring the consistency and reliability of the input data. Next, a feature extraction stage (Section 4.3) is performed to capture meaningful characteristics from the images. Finally, the extracted features are utilized for classification in Section 4.6 using a deep learning model to accurately distinguish between different diagnostic categories.

4.1 Dataset Description

The dataset used in this research was collected under the Indian Council of Medical Research (ICMR) Intermediate Extramural Project Grant (Project No. IIRPIG-2024-01-00483), titled “Development of Low-Cost Artificial Intelligent System for Prediction

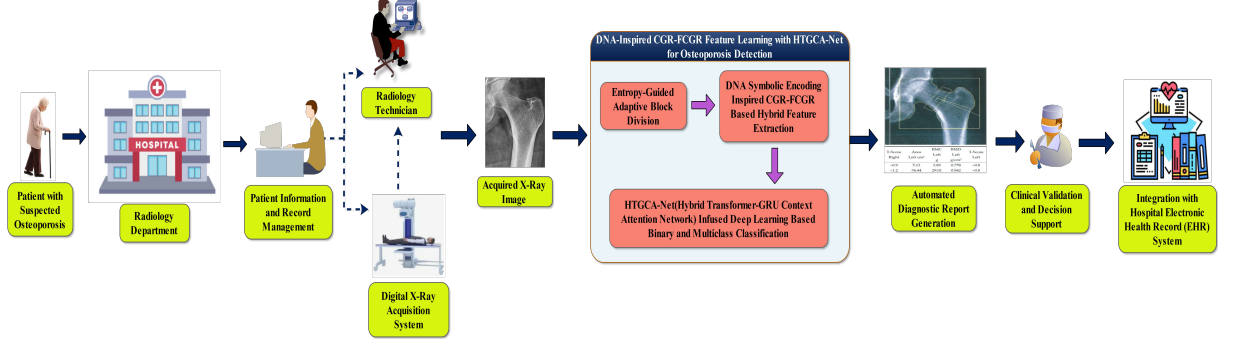


Fig. 1: Clinical workflow of the proposed osteoporosis classification system, from patient screening and X-ray acquisition to automated diagnosis and EHR integration.

of Bone Mineral Density (BMD) From Plain Radiographic Images in Osteoporosis Detection.” The dataset consists of plain radiographic images of the hip (proximal femur) obtained from multiple tertiary care institutes in India, including AIIMS Raipur, AIIMS Bhubaneswar, AIIMS Deoghar, and AIIMS Delhi. Image acquisition and data collection were conducted following an approved clinical protocol within the scope of the ICMR-funded project.

Ground-truth labels for each radiograph were established using Dual-Energy X-ray Absorptiometry (DXA) measurements of femoral bone mineral density. Diagnostic categorization was performed according to the World Health Organization (WHO) T-score criteria: Normal ($T \geq -1.0$), Osteopenia ($-2.5 < T < -1.0$), and Osteoporosis ($T \leq -2.5$). This labeling strategy ensures that the classification targets are clinically validated and consistent with established osteoporosis diagnostic standards.

The dataset contains a total of 1,678 hip radiographs. For multiclass classification, the dataset includes 713 Normal, 632 Osteopenia, and 333 Osteoporosis cases. For binary classification, Osteopenia cases were grouped with Normal to form the non-osteoporotic category, resulting in 1,345 Normal (non-osteoporotic) and 333 Osteoporosis images.

To maintain methodological rigor and eliminate the possibility of data leakage, all dataset partitions were performed strictly at the patient level. Images belonging to the same patient were assigned exclusively to a single subset. Multiple patient-wise data distributions were evaluated in this research to assess the robustness of the proposed framework. Specifically, four different training-testing splits were considered: 60:40, 70:30, 80:20, and 90:10. In each configuration, the specified proportion of patients was allocated for model development while the remaining patients were reserved for independent testing. Stratified sampling was applied to preserve the original class distribution across all partitions.

Regarding class imbalance, a limited augmentation strategy was applied only for the binary classification experiments to improve the representation of the minority Osteoporosis class within the training data. The augmentation included controlled geometric and photometric transformations such as small rotations ($\pm 10^\circ$), minor translations, slight zooming, horizontal flipping when anatomically appropriate, and

Table 4: Number of manually collected hip X-ray input images for Binary Class and Multiclass classification

	Binary Class		Multiclass		
	Normal	Osteoporosis	Normal	Osteopenia	Osteoporosis
No. of images	1345	333	713	632	333

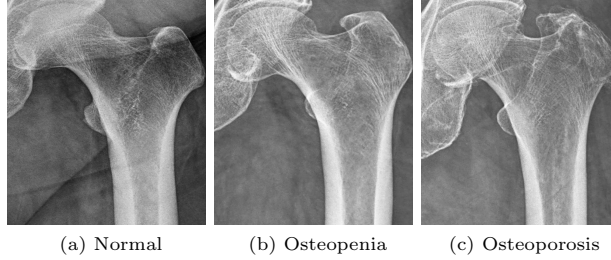


Fig. 2: Representative hip (proximal femur) radiographs from the ICMR dataset showing Normal, Osteopenia, and Osteoporosis categories.

mild brightness variation. These transformations were carefully chosen to maintain clinical realism while improving class balance. No augmentation was applied to the multiclass dataset, and the test sets in all experiments remained completely untouched to ensure unbiased performance evaluation.

4.2 Entropy-Guided Adaptive Block Division

An entropy-guided adaptive block division technique is used to improve spatial adaptivity in the localization of diagnostically relevant bone structures. By dynamically dividing X-ray images according to local textural complexity, as determined by Shannon entropy, this method improves the reliability of regional feature extraction. The proposed entropy-driven scheme adaptively subdivides regions exhibiting high structural irregularity or heterogeneous bone density key characteristics of osteoporotic degradation. As a result, the framework captures detailed morphological cues in complex areas while maintaining computational efficiency in homogeneous regions.

Initially, each preprocessed X-ray image is divided into a coarse 4×4 grid. The Shannon entropy (H) for each block is calculated as:

$$H = - \sum_{i=1}^N p_i \log_2(p_i) \quad (1)$$

where p_i represents the normalized histogram probability of the i^{th} gray-level intensity. A block is recursively subdivided into four (2×2) smaller sub-blocks if its entropy exceeds a predefined threshold ($H > 4.0$), thereby enhancing the spatial resolution in regions with complex bone texture and structural variation. Conversely, blocks with

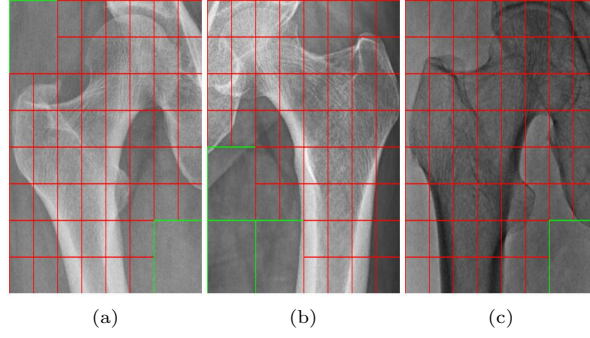


Fig. 3: Visual overview of Entropy Based Adaptive Block Division for (a) Normal, (b) Osteopenia, and (c) Osteoporosis images.

lower entropy values are retained, as they correspond to uniform or less informative bone regions. This adaptive process ensures focused feature extraction from diagnostically significant zones such as areas of trabecular thinning, cortical discontinuity, or bone porosity while minimizing redundant analysis.

As shown in Figure 3, red grid lines indicate high-entropy regions that are subdivided into finer partitions, whereas green lines represent low-entropy regions preserved at their original granularity. Subfigure (a) demonstrates the entropy map for a normal bone with uniform structure, (b) highlights moderate texture variation in osteopenia, and (c) reveals pronounced entropy heterogeneity in osteoporosis-affected bone. This entropy-driven decomposition enhances region-specific feature discriminability, enabling the hybrid feature extraction model to achieve improved diagnostic accuracy and interpretability.

The procedural steps for the entropy-based division are summarized in Algorithm 1.

This entropy-driven refinement process adaptively captures both global bone structure and localized textural irregularities, effectively emphasizing clinically relevant osteoporotic patterns. Figure 3 illustrates the entropy-guided block division, where high-entropy regions (shown in red) are automatically decomposed into finer sub-blocks for enhanced analysis.

4.3 DNA Symbolic Encoding of Grayscale Image

To enable the application of Chaos Game Representation (CGR) and Frequency Chaos Game Representation (FCGR), the grayscale X-ray image must first be transformed into a symbolic sequence. This is achieved through a DNA symbolic encoding scheme, which converts pixel intensities into nucleotide symbols. Such a representation allows the image data to be analyzed using sequence-based fractal and frequency analysis techniques while preserving the structural information embedded within the pixel intensity distribution.

Let the input grayscale image be represented as

Algorithm 1 Entropy-Based Adaptive Block Division

Require: Grayscale X-ray image I , initial grid size $G = 4$, entropy threshold $T = 4.0$

Ensure: Adaptive grid structure highlighting high-entropy regions

- 1: Resize image I to a uniform size (e.g., 256×256).
 - 2: Divide I into $G \times G$ coarse blocks.
 - 3: **for** each block $B_{i,j}$ in I **do**
 - 4: Compute Shannon entropy $H(B_{i,j}) = -\sum p_i \log_2(p_i)$.
 - 5: **if** $H(B_{i,j}) > T$ **then**
 - 6: Subdivide $B_{i,j}$ into 2×2 sub-blocks.
 - 7: Mark sub-blocks with red boundary (high entropy).
 - 8: **else**
 - 9: Retain $B_{i,j}$ without subdivision.
 - 10: Mark block with green boundary (low entropy).
 - 11: **end if**
 - 12: **end for**
 - 13: Visualize adaptive grid overlay on I .
-

$$I = \{I(i, j)\}, \quad I(i, j) \in [0, 255], \quad 1 \leq i \leq M, \quad 1 \leq j \leq N \quad (2)$$

where M and N denote the height and width of the image, respectively, and $I(i, j)$ represents the intensity value of the pixel located at coordinates (i, j) . Each pixel value is encoded using an 8-bit binary representation, which provides a compact and standardized digital representation of grayscale intensity.

For a given pixel value $p = I(i, j)$, its binary representation is defined as

$$B = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8) \quad (3)$$

where $b_k \in \{0, 1\}$ for $k = 1, 2, \dots, 8$. The binary sequence is then partitioned into four non-overlapping 2-bit segments, defined as

$$P_k = (b_{2k-1}, b_{2k}), \quad k = 1, 2, 3, 4 \quad (4)$$

Each binary pair P_k is subsequently mapped to a nucleotide symbol according to a predefined binary-to-DNA encoding rule 1 from Table 5. The mapping function $f(\cdot)$ is defined as

$$f(P_k) = \begin{cases} A, & P_k = 00 \\ G, & P_k = 01 \\ C, & P_k = 10 \\ T, & P_k = 11 \end{cases} \quad (5)$$

where A , T , G , and C denote the nucleotide symbols adenine, thymine, guanine, and cytosine, respectively. This transformation converts each pixel into four symbolic elements, thereby generating a DNA-like symbolic representation of the image.

Formally, the symbolic encoding process produces a sequence

$$S = \{s_1, s_2, \dots, s_L\} \quad (6)$$

DNA Nucleotides				
Original	A	G	T	C
Binary	00	01	11	10
Complement	11	10	00	01
Complementary DNA	T	C	A	G
DNA ENCODING RULES				
Rule 1	00	01	11	10
Rule 2	11	10	00	01
Rule 3	10	11	01	00
Rule 4	01	00	10	11
Rule 5	00	11	01	10
Rule 6	11	00	10	01
Rule 7	10	01	00	11
Rule 8	01	10	11	00

Table 5: Watson-Crick rule DNA Encoding Rules and it's Complementary Pairing

where $s_L \in \{A, T, G, C\}$ and the sequence length is given by

$$L = 4 \times M \times N \quad (7)$$

since each pixel contributes four nucleotide symbols.

The transformation from pixel intensities to nucleotide symbols preserves the local intensity transitions within the image while enabling sequence-based analytical techniques such as CGR and FCGR. Consequently, the resulting symbolic sequence provides a compact representation that captures both local pixel variations and underlying structural patterns of the X-ray image. Thus, a single grayscale pixel is transformed into a four-symbol nucleotide subsequence. Repeating this process for all pixels in the image into a complete DNA symbolic sequence representing the entire image.

4.4 CGR and FCGR inspired Spatial and frequency domain based Hybrid Feature Extraction

4.4.1 Chaos Game Representation (CGR)-Based Spatial domain Feature Extraction

The *Chaos Game Representation (CGR)*, initially introduced by Jeffrey[16][15], is a geometric mapping technique that encodes symbolic sequences into a two-dimensional coordinate space. In its classical genomic formulation, a symbolic nucleotide sequence composed of adenine (A), thymine (T), cytosine (C), and guanine (G) is mapped into a unit square, producing a self-similar fractal pattern that reflects the statistical organization of the sequence. Regular biological sequences generate structured fractal attractors, whereas random or uncorrelated sequences produce nearly uniform scatter distributions. This property makes CGR an effective approach for translating symbolic

Algorithm 2 DNA Symbolic Encoding for Image-to-Sequence Transformation

Require: Grayscale image $I \in \mathbb{R}^{M \times N}$, where $I(i, j) \in [0, 255]$ represents the intensity of the pixel located at coordinates (i, j) , and M and N denote the height and width of the image, respectively.

Ensure: DNA symbolic sequence $S = \{s_1, s_2, \dots, s_L\}$ where $s_l \in \{A, T, G, C\}$ and $L = 4MN$.

```
1: Initialize an empty symbolic sequence  $S \leftarrow \emptyset$ 
2: for  $i = 1$  to  $M$  do
3:   for  $j = 1$  to  $N$  do
4:      $p \leftarrow I(i, j)$  ▷ Pixel intensity value
5:     Convert  $p$  to its 8-bit binary representation
6:      $B = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8)$ ,  $b_k \in \{0, 1\}$ 
7:     Partition  $B$  into four non-overlapping binary pairs:
8:      $P_1 = (b_1b_2)$ ,  $P_2 = (b_3b_4)$ ,  $P_3 = (b_5b_6)$ ,  $P_4 = (b_7b_8)$ 
9:     for each binary pair  $P_k$ ,  $k \in \{1, 2, 3, 4\}$  do
10:      if  $P_k = 00$  then
11:        Append symbol  $A$  to sequence  $S$ 
12:      else if  $P_k = 01$  then
13:        Append symbol  $G$  to sequence  $S$ 
14:      else if  $P_k = 10$  then
15:        Append symbol  $C$  to sequence  $S$ 
16:      else if  $P_k = 11$  then
17:        Append symbol  $T$  to sequence  $S$ 
18:      end if
19:    end for
20:  end for
21: end for
22: return  $S$ 
```

sequences into spatially analyzable patterns while preserving both compositional and sequential correlations.

In the proposed framework, the CGR methodology is adapted to grayscale medical X-ray images through a DNA-inspired symbolic encoding process. Prior to applying CGR, each grayscale image undergoes an entropy-guided adaptive block division, allowing regions with higher textural complexity to be analyzed with finer granularity while maintaining computational efficiency in homogeneous areas.

Within each block, pixel intensity values are transformed into a symbolic nucleotide sequence using the proposed DNA symbolic encoding technique. Let the grayscale image be denoted as

$$I = \{I(i, j)\}, \quad I(i, j) \in [0, 255]$$

where $I(i, j)$ represents the intensity value at pixel location (i, j) .

Each pixel value is first converted into its 8-bit binary representation

$$B = (b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8)$$

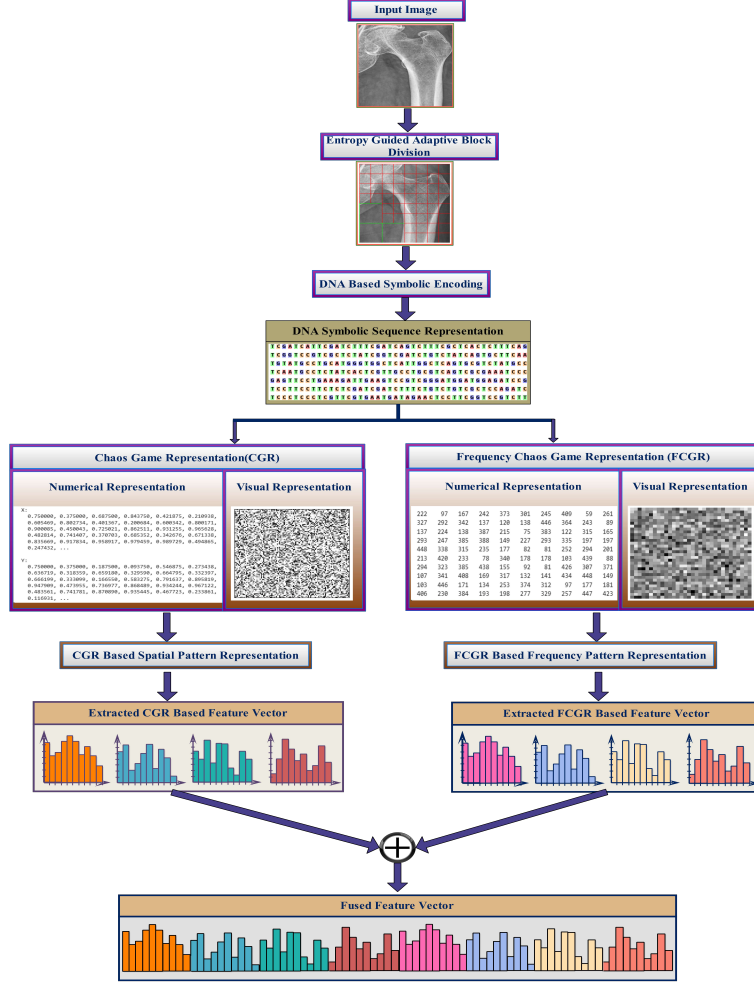


Fig. 4: Overview of the Proposed Feature Extraction Framework for Osteoporosis Detection Using CGR and FCGR Representations

where $b_k \in \{0,1\}$. The binary sequence is then partitioned into four non-overlapping pairs

$$(b_1b_2), (b_3b_4), (b_5b_6), (b_7b_8)$$

Each binary pair is mapped to a nucleotide symbol according to the encoding rule 1 (Table 5)

$$00 \rightarrow A, \quad 01 \rightarrow G, \quad 10 \rightarrow C, \quad 11 \rightarrow T$$

Through this transformation, the grayscale image block is converted into a symbolic nucleotide sequence

$$S = \{s_1, s_2, \dots, s_L\}, \quad s_n \in \{A, T, G, C\}$$

where L denotes the total length of the generated sequence.

Once the symbolic sequence is obtained, the standard CGR algorithm is applied. The CGR process begins from an initial starting point located at the center of the unit square

$$(x_0, y_0) = (0.5, 0.5)$$

For each successive symbol s_n in the sequence, a new coordinate (x_n, y_n) is computed as the midpoint between the previous coordinate (x_{n-1}, y_{n-1}) and the vertex associated with symbol s_n . The iterative transformation is defined as

$$\begin{aligned} x_n &= \frac{x_{n-1} + x_{s_n}}{2} \\ y_n &= \frac{y_{n-1} + y_{s_n}}{2} \end{aligned}$$

where

- (x_{n-1}, y_{n-1}) denotes the coordinate generated at the previous iteration
- (x_{s_n}, y_{s_n}) denotes the vertex coordinate corresponding to the nucleotide symbol s_n
- (x_n, y_n) represents the newly generated CGR coordinate

By recursively applying this transformation for all symbols in the sequence, a set of CGR coordinates is obtained

$$P = \{(x_1, y_1), (x_2, y_2), \dots, (x_L, y_L)\}$$

which forms a spatial distribution reflecting the recurrence characteristics and structural organization of the encoded sequence.

To extract discriminative spatial descriptors, the CGR coordinate space is discretized into a two-dimensional grid of size $R \times R$, producing a CGR density matrix

$$C(i, j), \quad 1 \leq i, j \leq R$$

where $C(i, j)$ represents the number of CGR points falling within the grid cell located at position (i, j) .

To reduce dynamic range variations and improve numerical stability, logarithmic normalization is applied

$$C_{norm}(i, j) = \log(1 + C(i, j))$$

where $C_{norm}(i, j)$ denotes the normalized CGR value.

The normalized matrix is then flattened into a one-dimensional feature vector

$$\mathbf{f}_{CGR} \in \mathbb{R}^{R \times R}$$

which represents the CGR-based feature descriptor extracted from each image block. Figure 5 illustrates the Chaos Game Representation (CGR) process applied

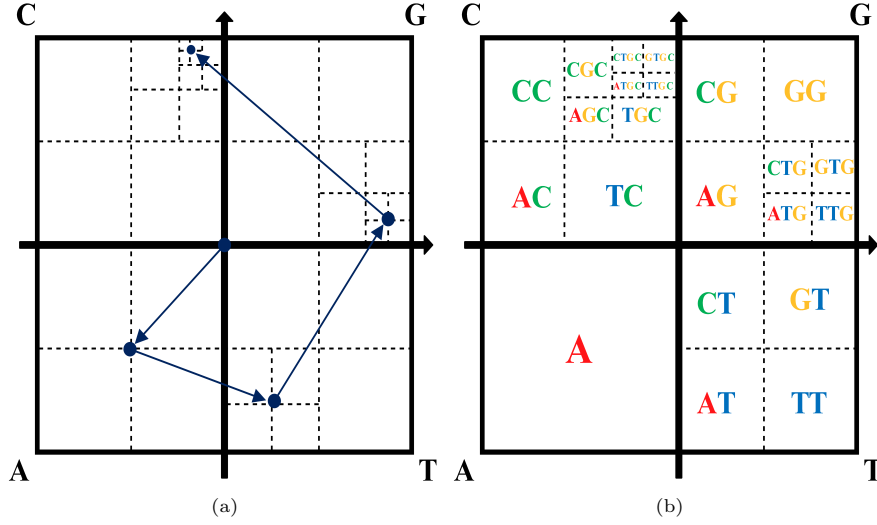


Fig. 5: Chaos Game Representation (CGR) of DNA-encoded sequences. (a) iterative point generation illustrating the sequence-dependent trajectory. (b) symbolic mapping within the CGR coordinate space defined by vertices A, T, G, and C.

to the DNA-encoded symbolic sequence. The CGR space is defined by assigning the four vertices of a unit square to the symbols A, C, G, and T, enabling a structured mapping of the sequence. Starting from the center, each subsequent point is iteratively generated toward the vertex corresponding to the current symbol, forming a trajectory that reflects the sequence order. The resulting spatial distribution of points captures the underlying structural patterns of the encoded data, which are later utilized for feature extraction.

From each CGR representation, several categories of features are implicitly encoded:

1. **Symbolic frequency statistics** - capturing the occurrence distribution of encoded nucleotide symbols.
2. **Spatial fractal descriptors** - representing self-similar spatial structures within the CGR pattern.
3. **Entropy and energy measures** - describing local randomness and structural irregularity of bone texture.
4. **Uniformity and variance measures** - quantifying spatial regularity and density fluctuations within image regions.

Collectively, these descriptors provide a biologically inspired multi-level representation that CGR primarily captures global structural patterns while implicitly encoding local relationships through spatial point distributions. The integration of entropy-guided adaptive partitioning, DNA-based symbolic encoding, and CGR-based

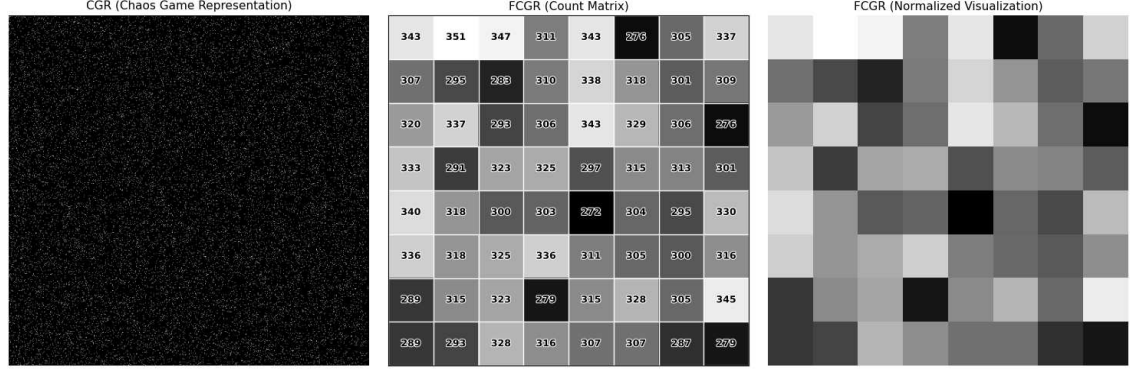


Fig. 6: CGR-FCGR feature extraction from X-ray images. Left: CGR representation capturing spatial patterns; middle: FCGR count matrix; right: normalized FCGR highlighting global structural variations.

fractal feature extraction enables the proposed framework to effectively characterize structural variations in bone density patterns from X-ray images.

4.4.2 Frequency Matrix Chaos Game Representation (FCGR)-Based Frequency Domain Feature Extraction

The *Frequency Chaos Game Representation (FCGR)*[15] is an advanced extension of the traditional CGR framework that quantifies the frequency distribution of nucleotide subsequences, or *k-mers*, within a symbolic sequence. While the classical CGR provides a fractal-based geometric visualization of sequence composition and recurrence, FCGR transforms this representation into a statistical matrix capturing the normalized occurrence of all possible *k*-length nucleotide combinations. This representation encodes both local and global dependencies within the symbolic sequence, resulting in a scale-invariant and information-dense descriptor.

In the proposed framework, the symbolic nucleotide sequence used for FCGR computation is obtained through the DNA symbolic encoding procedure described in the previous subsection. Through this encoding process, grayscale X-ray image blocks are transformed into a sequence

$$S = \{s_1, s_2, s_3, \dots, s_L\}$$

where

- $s_i \in \{A, T, G, C\}$ denotes the i^{th} nucleotide symbol in the sequence
- L represents the total length of the symbolic sequence

A *k-mer* is defined as a contiguous subsequence of length k extracted from S :

$$K_i = (s_i, s_{i+1}, \dots, s_{i+k-1})$$

where $i = 1, 2, \dots, L - k + 1$.

The total number of possible nucleotide combinations of length k is

$$N_k = 4^k$$

since each position of the subsequence can take one of the four symbols $\{A, T, G, C\}$.

For each encoded image block, the FCGR computation is performed by counting the frequency of all possible k -mer combinations. For example, when $k = 3$, the set of possible combinations includes

$$\{AAA, AAC, AAG, \dots, TTT\}.$$

Let the frequency of a particular k -mer pattern ω in the sequence S be defined as

$$F(\omega) = \sum_{i=1}^{L-k+1} \delta(K_i, \omega)$$

where

- $F(\omega)$ denotes the number of occurrences of the k -mer ω
- $\delta(K_i, \omega)$ is an indicator function defined as

$$\delta(K_i, \omega) = \begin{cases} 1 & \text{if } K_i = \omega \\ 0 & \text{otherwise} \end{cases}$$

The FCGR matrix is constructed by assigning each possible k -mer to a unique cell in a two-dimensional matrix:

$$M \in \mathbb{R}^{2^k \times 2^k}$$

where

$$M(i, j) = F(w_{i,j})$$

and $w_{i,j}$ denotes the nucleotide subsequence associated with the matrix coordinate (i, j) .

To ensure comparability across sequences of different lengths, the frequency values are normalized to obtain a probabilistic representation:

$$P(i, j) = \frac{M(i, j)}{\sum_{i=1}^{2^k} \sum_{j=1}^{2^k} M(i, j)}$$

where

- $P(i, j)$ represents the normalized probability of the k -mer corresponding to position (i, j)
- $\sum M(i, j)$ represents the total number of extracted k -mers in the sequence

In addition to k -mer statistics, symbol-level frequency features are computed by measuring the normalized occurrence of each nucleotide symbol:

$$P(s) = \frac{n_s}{L}$$

where

- n_s denotes the number of occurrences of symbol s
- $s \in \{A, T, G, C\}$
- L represents the total length of the symbolic sequence

The FCGR matrix is subsequently flattened to obtain the FCGR feature vector

$$\mathbf{f}_{FCGR} \in \mathbb{R}^{(2^k \times 2^k)}$$

which encodes the statistical distribution of nucleotide subsequences within the symbolic representation of the image.

From the resulting FCGR map, several categories of features are derived:

1. **Symbolic frequency features** - describing the global occurrence statistics of encoded nucleotide symbols that reflect grayscale intensity variations.
2. **K-mer compositional patterns** - capturing repetitive structural motifs and local dependencies through the frequency analysis of ordered nucleotide subsequences.
3. **Entropy-driven texture descriptors** - integrating adaptive block division to emphasize structural heterogeneity and trabecular irregularity.
4. **Normalized probabilistic measures** - ensuring comparability and stability of features across variable region sizes and scales.

The fusion of entropy-based spatial adaptivity and FCGR-based symbolic frequency modeling yields a biologically inspired statistical abstraction of bone microstructure. This encoding captures subtle morphological variations, density asymmetries, and irregular bone patterns that are often imperceptible to conventional pixel-level analysis. Consequently, the FCGR-derived feature representation provides a robust and discriminative basis for osteoporosis detection and grading from medical X-ray images.

4.5 Feature Vector Construction

The dimensionality of the extracted feature vectors is determined by the discretization parameters of the CGR and FCGR representations.

For the CGR representation, the generated coordinate distribution is discretized into a two-dimensional density matrix of size $R \times R$. Each grid cell records the number of CGR points falling within that spatial region. The resulting matrix therefore contains

$$d_{CGR} = R \times R$$

spatial descriptors corresponding to the CGR density distribution.

In the proposed framework, the CGR matrix is further processed to compute additional spatial statistics including density distribution, fractal spatial characteristics, and statistical descriptors extracted from each grid region. These operations produce a total of

$$d_{CGR} = 66\,560$$

CGR-based spatial features.

For the FCGR representation, the symbolic sequence is analyzed using k -mer frequency statistics. For a nucleotide alphabet $\{A, T, G, C\}$, the number of possible k -length subsequences is given by

$$N_k = 4^k$$

Each possible k -mer is assigned to a unique cell in the FCGR matrix, resulting in a matrix of size

$$2^k \times 2^k$$

where each cell represents the frequency of a specific nucleotide subsequence.

After normalization and feature extraction, the FCGR representation produces

$$d_{FCGR} = 5120$$

frequency-based descriptors capturing k -mer composition patterns and symbolic frequency statistics of the encoded image sequence.

To exploit the complementary characteristics of spatial and frequency representations, the CGR and FCGR feature vectors are concatenated to form a unified feature representation:

$$F = [f_{CGR}, f_{FCGR}]$$

where

$$F \in \mathbb{R}^d$$

and the total feature dimensionality is

$$d = d_{CGR} + d_{FCGR}$$

$$d = 66560 + 5120 = 71680$$

Thus, the proposed hybrid feature extraction framework produces a high-dimensional descriptor consisting of 71,680 features that integrate spatial fractal characteristics from CGR and statistical frequency patterns from FCGR. Although the resulting feature space is high-dimensional, the hybrid deep learning classifier is capable of learning compact and discriminative representations from the fused feature vectors.

4.6 Hybrid Deep Learning Based Binary class and Multiclass Classification Architecture

Effectively learn complex relationships among the extracted CGR and FCGR feature representations, a hybrid deep learning architecture **HTGCA-Net**(Hybrid Transformer-GRU Context Attention Network) is designed by integrating convolutional, transformer-based attention, recurrent modeling, and context-based attention mechanisms. The proposed architecture combines the strengths of Convolutional Neural Networks (CNN), Transformer encoders, Bidirectional Gated Recurrent Units (BiGRU), and an attention-with-context mechanism to capture both local and global dependencies within the feature sequences.

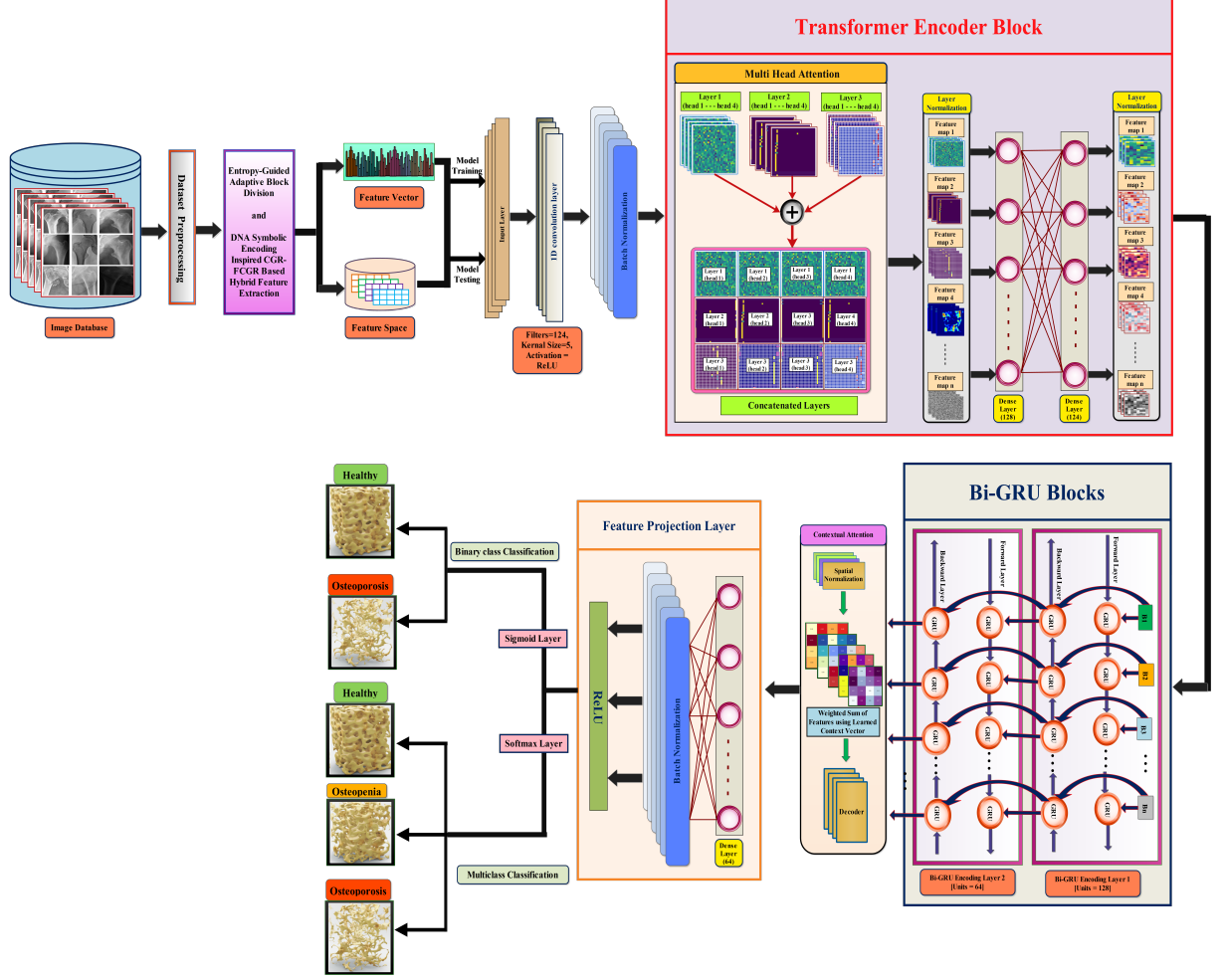


Fig. 7: Detail Architecture of Hybrid Transformer-GRU Context Attention Network(HTGCA-Net) for Classification process

4.6.1 Input Representation

Let the extracted feature matrix be denoted as

$$X \in \mathbb{R}^{T \times F}$$

where T represents the number of feature timesteps (or feature sequence length) and F represents the number of features at each timestep. The input sequence is fed into the hybrid architecture for hierarchical feature learning.

4.6.2 Convolutional Feature Extraction

A one-dimensional convolution layer is first applied to capture local dependencies among neighboring features. The convolution operation is expressed as:

$$h_i = \sigma \left(\sum_{k=0}^{K-1} W_k \cdot X_{i+k} + b \right)$$

where:

- h_i denotes the output feature at position i
- K is the kernel size
- W_k represents the convolutional kernel weights
- b is the bias term
- $\sigma(\cdot)$ denotes the activation function (ReLU)

This layer enables the model to extract localized feature patterns and correlations among adjacent descriptors.

4.6.3 Transformer Encoder for Global Dependency Modeling

To capture long-range dependencies among extracted features, a Transformer encoder block is employed. The core component of the Transformer is the multi-head self-attention mechanism.

For a given input sequence X , query, key, and value matrices are computed as:

$$Q = XW_Q, \quad K = XW_K, \quad V = XW_V$$

where W_Q , W_K , and W_V are learnable projection matrices.

The scaled dot-product attention is then computed as:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

where:

- d_k represents the dimensionality of the key vectors
- QK^T measures similarity between query and key vectors
- the softmax function generates normalized attention weights

Multi-head attention further enhances representation learning by computing attention across multiple subspaces:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O$$

where h is the number of attention heads and W_O is the output projection matrix.

This mechanism allows the network to model global feature interactions within the sequence.

4.6.4 Bidirectional GRU for Sequential Feature Modeling

After capturing global dependencies, sequential relationships among features are modeled using Bidirectional Gated Recurrent Units (BiGRU). The GRU unit updates its hidden state as follows:

$$z_t = \sigma(W_z x_t + U_z h_{t-1})$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1})$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1}))$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t$$

where:

- x_t is the input at timestep t
- h_t is the hidden state
- z_t denotes the update gate
- r_t denotes the reset gate
- $\odot$ represents element-wise multiplication
- W and U are learnable weight matrices

The bidirectional configuration processes the sequence in both forward and backward directions, allowing the network to capture contextual dependencies from both past and future feature representations.

4.6.5 Attention with Context Mechanism

To emphasize the most discriminative features, an attention-with-context mechanism is applied. Given the hidden states h_i produced by the BiGRU layer, an attention score is computed as:

$$u_i = \tanh(W_a h_i + b_a)$$

The importance weight of each feature is then calculated using a context vector u_c :

$$\alpha_i = \frac{\exp(u_i^T u_c)}{\sum_j \exp(u_j^T u_c)}$$

where:

- u_i represents the hidden representation of timestep i
- u_c is the trainable context vector
- α_i denotes the attention weight assigned to feature i

The final context-aware feature representation is computed as:

$$v = \sum_i \alpha_i h_i$$

This mechanism enables the model to focus on the most informative features relevant for osteoporosis classification.

4.6.6 Classification Layer

The aggregated context vector v is passed through fully connected layers for classification. For multiclass classification (Normal, Osteopenia, Osteoporosis), the softmax function is used:

$$P(y = c|v) = \frac{\exp(W_c v + b_c)}{\sum_{k=1}^C \exp(W_k v + b_k)}$$

where:

- C denotes the number of classes
- W_c and b_c are learnable parameters

For binary classification (Normal vs. Osteoporosis), the sigmoid function is applied:

$$P(y = 1|v) = \frac{1}{1 + \exp(-(Wv + b))}$$

This hybrid architecture enables the model to capture local patterns, global feature interactions, sequential dependencies, and feature importance simultaneously, resulting in improved classification performance for both binary and multiclass osteoporosis detection tasks.

5 Results and Discussion

To assess the effectiveness and robustness of the proposed *HTGCA-Net* (Hybrid Transformer-GRU Context Attention Network), a comprehensive set of experiments was performed. This section outlines the experimental protocol, including the dataset characteristics, evaluation metrics, and implementation details used for performance analysis. The experiments were conducted on a manually curated dataset of hip X-ray images for osteoporosis assessment. The evaluation is structured into two classification settings: (i) binary classification, where the task is to distinguish between normal and osteoporotic cases, and (ii) multi-class classification, involving Normal, Osteopenia, and Osteoporosis categories. For each setting, multiple data split configurations were considered to examine the consistency and generalization capability of the proposed model. Performance is analyzed using standard quantitative metrics along with feature-specific evaluations, enabling a comprehensive understanding of the model behavior under different experimental conditions. Detailed results and discussions are presented in the subsequent subsections.

5.1 Experimental Setup

The proposed hybrid model was implemented using Python 3.10 and TensorFlow 2.17, and which was run on a system with an Intel Core i9 processor, 64 GB of RAM, and an NVIDIA RTX 3090 GPU with 24 GB of VRAM. Several stratified train-test split configurations, such as 60:40, 70:30, 80:20, and 90:10, were used to divide the dataset for model construction in order to maintain class distribution. This multi-split evaluation approach was used to methodically examine the suggested

model’s resilience and capacity for generalization under various training data ratios. In particular, higher ratios (e.g., 90:10) evaluate the model’s ability to leverage bigger training sets, whereas lower ratios (e.g., 60:40) allow evaluation under restricted data availability. Because it offers a fair trade-off between training sufficiency and testing reliability and is frequently used in medical imaging research, the 80:20 split was chosen among these configurations to report the primary results.

The input features were structured as sequential representations and processed using a hybrid deep learning architecture. Initially, a Conv1D layer with 124 filters and kernel size of 5 was employed for local feature extraction, followed by batch normalization and dropout regularization. The feature maps were then passed through two stacked Transformer encoder blocks, each consisting of multi-head self-attention with 4 attention heads and a key dimension of 64, along with a feed-forward network of dimension 128 and dropout rate of 0.1. These blocks facilitate modeling of global contextual dependencies within the feature space. Subsequently, two bidirectional GRU layers with 128 and 64 hidden units, respectively, were utilized to capture sequential dependencies. An attention-with-context mechanism was applied to aggregate the most informative features into a compact representation. This was followed by a fully connected layer with 64 neurons and batch normalization before the final classification layer. The model was optimized using the AdamW optimizer with an initial learning rate of 1×10^{-4} and trained using sparse categorical cross-entropy loss. Training was done for 30 epochs with a batch size of 16. A ReduceLROnPlateau scheduler with a reduction factor of 0.2 and patience of five epochs was used to enhance convergence. To reduce overfitting, dropout layers with rates of 0.3 and 0.2 were added at various points. In order to prevent data leaking and guarantee that photos from the same patient do not show in both training and testing sets, a patient-wise data splitting approach was also included. This improved the evaluation’s generality and reliability.

5.2 Performance for Binary Class Classification

The binary classification task was designed to distinguish between Normal and Osteoporosis cases. The stability and generalization capabilities of the suggested model were assessed under many stratified dataset split ratios (60:40, 70:30, 80:20, and 90:10) using different quantities of training data. Three methods were examined for feature extraction. First, spatial distribution patterns, fractal features, and structural self-similarity arising from the symbolic representation of grayscale intensity fluctuations were captured using CGR-based features. Second, FCGR-based features were used, which represents global statistical dependencies inside the symbolic sequence and convey frequency-based information through k -mer distributions. Third, in order to provide a more thorough description of trabecular bone patterns, a hybrid feature representation (CGR + FCGR) that combines both spatial and frequency-domain properties was taken into consideration. This hybrid strategy was used to take use of the complementing qualities of FCGR and CGR characteristics, improving the model’s capacity for discrimination. To ensure a fair and trustworthy evaluation of classification performance across many evaluation criteria, the performance of each feature representation was assessed using common quantitative measures, such as Accuracy, Precision, Recall (Sensitivity), Specificity, and F1-score.

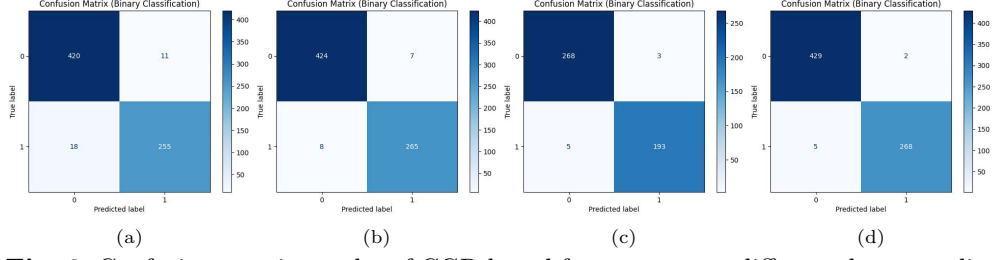


Fig. 8: Confusion matrix results of CGR based features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

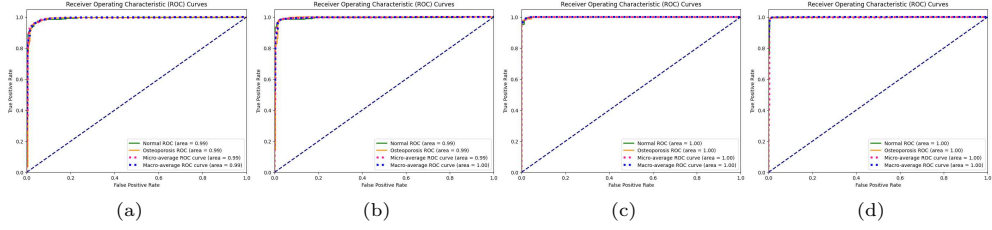


Fig. 9: ROC Curve results of CGR based features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

In addition to these metrics, confusion matrices were generated for each method to observe true positive, true negative, false positive, and false negative classification counts, offering insight into the model's misclassification patterns. ROC (Receiver Operating Characteristic) curves were plotted to demonstrate the trade-off between sensitivity and specificity and to assess the discriminative ability of each feature set. The experimental results indicate that the hybrid feature representation provides more consistent and balanced performance compared to individual feature sets. This is explained by its capacity to combine frequency-based statistical features from FCGR with spatial structural data from CGR to provide a more thorough description of trabecular bone structures. Although the performance improvements are marginal in certain cases, the hybrid approach demonstrates stable behavior across different evaluation metrics. In order to reduce both false positives and false negatives in medical diagnosis, the model achieves dependable classification accuracy while maintaining a well-balanced trade-off between sensitivity and specificity. Higher training ratios (e.g., 80:20 and 90:10) significantly enhance stability, demonstrating the model's robustness and its use in assisting with clinical osteoporosis assessment.

In certain experimental settings, comparable accuracy values were observed across CGR, FCGR, and hybrid feature representations. This behavior can be attributed to the presence of sufficiently discriminative information within individual feature sets, particularly in relatively well-separated data scenarios. Nevertheless, the hybrid

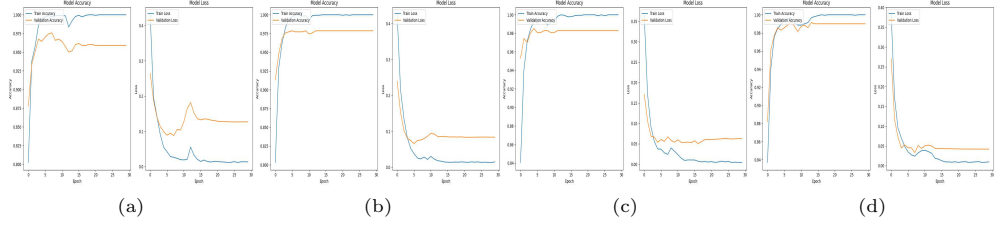


Fig. 10: Loss and Accuracy graph of CGR based features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

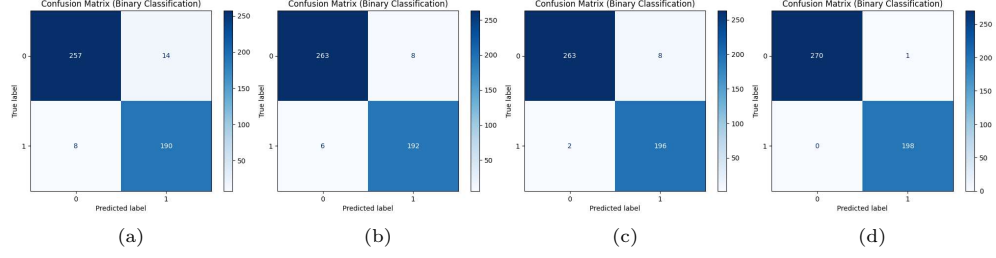


Fig. 11: Confusion matrix results of FCGR based features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

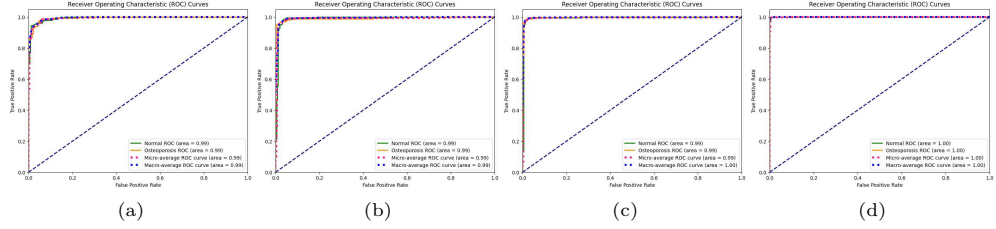


Fig. 12: ROC Curve results of FCGR based features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

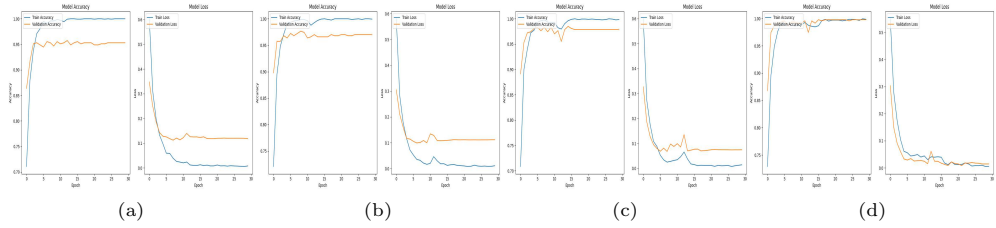


Fig. 13: Loss and Accuracy graph of FCGR based features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

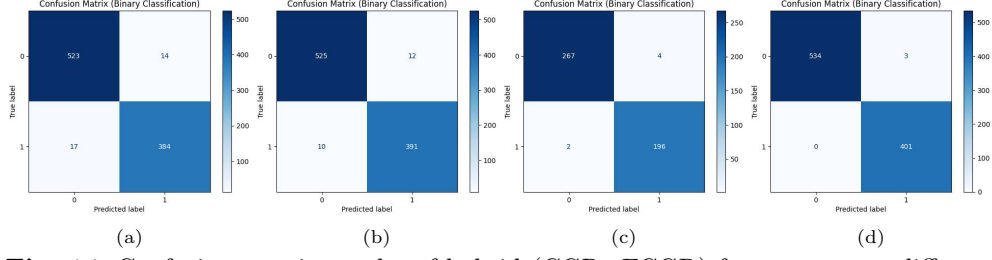


Fig. 14: Confusion matrix results of hybrid (CGR+FCGR) features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

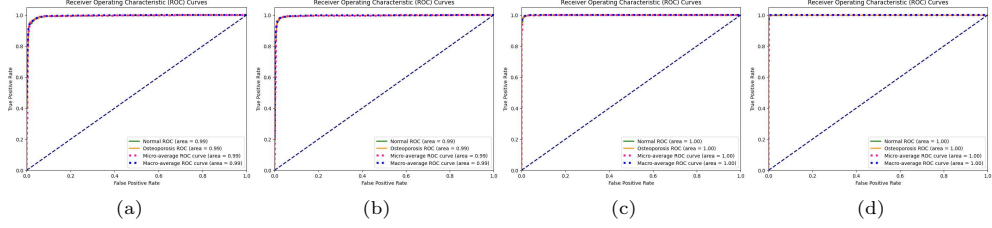


Fig. 15: ROC Curve results of hybrid (CGR+FCGR) features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

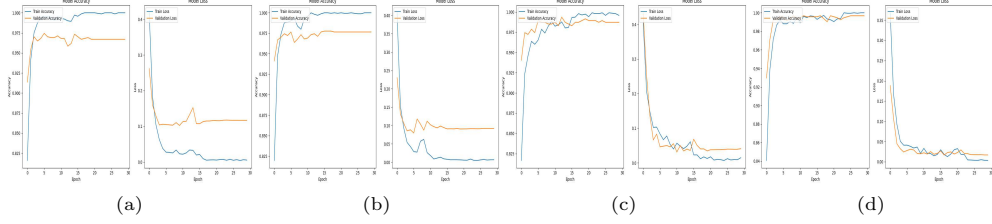


Fig. 16: Loss and Accuracy graph of hybrid (CGR+FCGR) features across different dataset splits in Binary Class classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

representation offers a more comprehensive characterization by integrating both spatial and frequency-based information, which contributes to improved stability and a more balanced trade-off between sensitivity and specificity. Overall, the hybrid approach demonstrates consistent or improved performance across most experimental configurations.

Table 6: Performance comparison of Binary-Class Classification across Different Methods (80-20 split)

Class	CGR Based Features			FCGR Based Features			Hybrid (CGR + FCGR) Features		
	P _r	Recall	F _s	P _r	Recall	F _s	P _r	Recall	F _s
Normal	0.98	0.97	0.97	0.99	0.97	0.98	0.99	0.99	0.99
Osteoporosis	0.98	0.97	0.97	0.96	0.98	0.98	0.98	0.99	0.98
Macro Average	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99
Weighted Average	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99

5.3 Performance for Multiclass Class Classification

For multiclass classification, the proposed framework was designed to distinguish among three diagnostic categories: Normal, Osteopenia, and Osteoporosis. Multiple experiments were conducted to evaluate the robustness of the proposed HTGCA-Net architecture under varying data availability using multiple stratified train-test split ratios (60:40, 70:30, 80:20, and 90:10).

Three feature representations were considered: (i) CGR-based features capturing spatial and fractal characteristics, (ii) FCGR-based features encoding frequency-driven statistical patterns, and (iii) a hybrid representation integrating both CGR and FCGR features. This comparative setup enables a systematic assessment of the contribution of individual and combined feature modalities.

Standard metrics, such as overall accuracy, class-wise precision, recall, and F1-score, as well as macro-averaged sensitivity and specificity to take possible class imbalance into consideration, were used to assess the model’s performance. ROC curves were used to evaluate separability between the three categories, and confusion matrices were examined to look at class-wise prediction behavior. In addition, training and validation accuracy-loss curves were monitored to ensure stable convergence and generalization of the model.

The results indicate that CGR and FCGR feature representations achieve comparable performance, whereas the hybrid feature representation provides more consistent and balanced results across different split configurations. This improvement can be attributed to the complementary integration of spatial structural information from CGR and frequency-based statistical characteristics from FCGR, which together offer a more comprehensive description of trabecular bone patterns. Additionally, the hybrid representation shows enhanced ability to differentiate between Osteopenia and Osteoporosis classes, where subtle structural variations are present. Overall, these findings suggest that the proposed feature fusion strategy enhances the discriminative capacity and stability of the model, making it suitable for multiclass osteoporosis classification tasks.

6 Feature Ablation Analysis

To evaluate the contribution of individual components in the proposed hybrid architecture, an ablation study was performed by systematically removing key modules while maintaining identical training settings and hyperparameters. The complete model consists of a Conv1D-based feature extraction layer followed by Transformer encoder

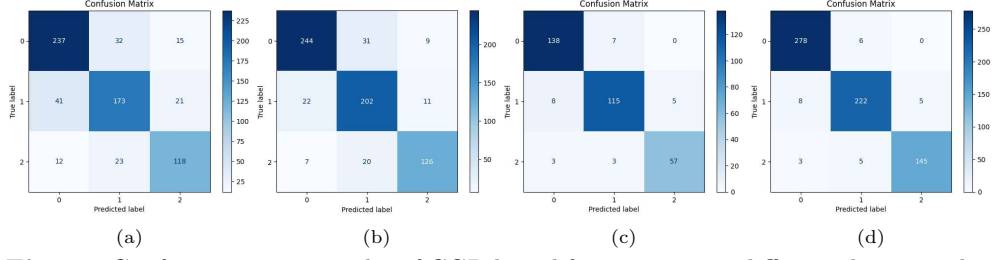


Fig. 17: Confusion matrix results of CGR based features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

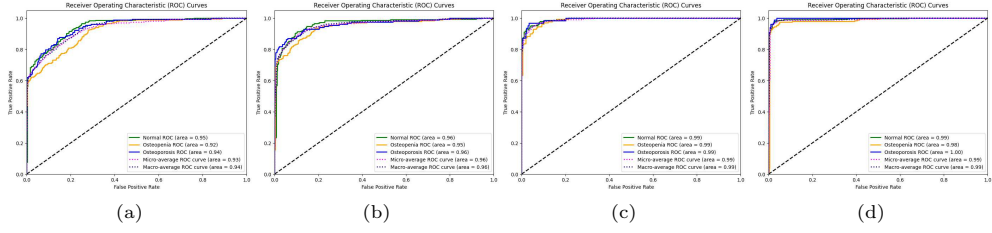


Fig. 18: ROC Curve results of CGR based features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

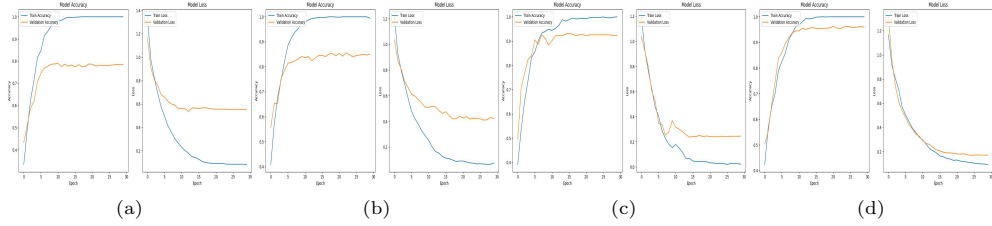


Fig. 19: Loss and Accuracy graph of CGR based features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

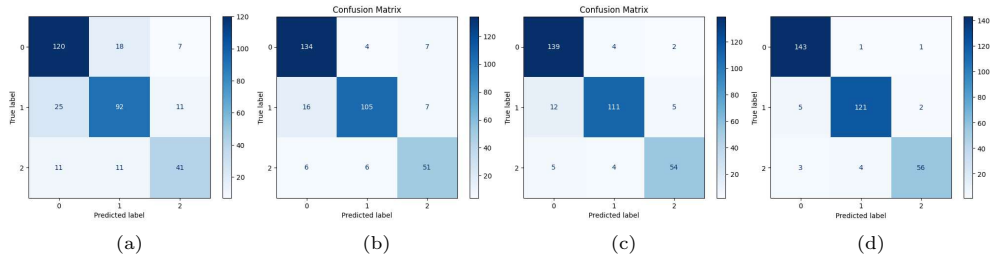


Fig. 20: Confusion matrix results of FCGR based features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

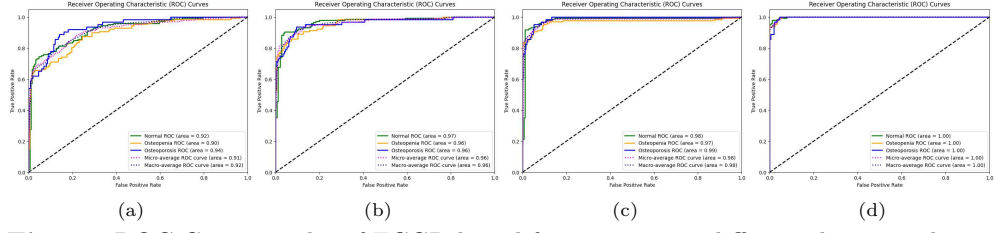


Fig. 21: ROC Curve results of FCGR based features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

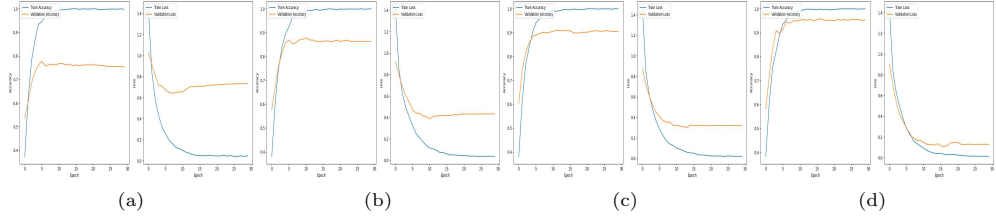


Fig. 22: Loss and Accuracy graph of FCGR based features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

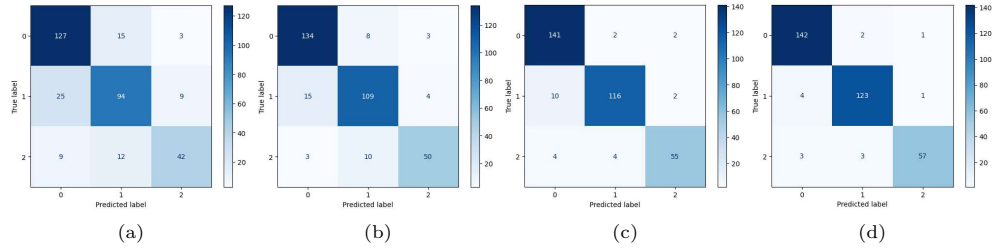


Fig. 23: Loss and Accuracy graph of hybrid (CGR+FCGR) features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

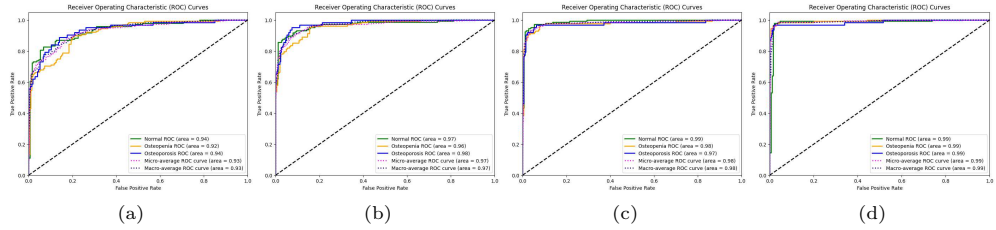


Fig. 24: ROC Curve results of hybrid (CGR+FCGR) features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

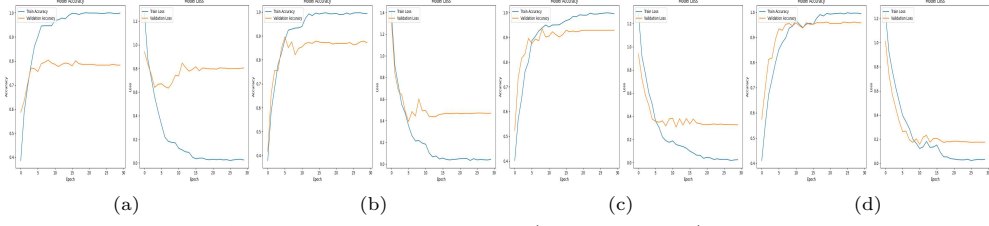


Fig. 25: Loss and Accuracy graph of hybrid (CGR+FCGR) features across different dataset splits in Multiclass classification: (a) 60:40, (b) 70:30, (c) 80:20, and (d) 90:10.

Table 7: Performance comparison of Multiclass Classification across Different Methods (80-20 split)

Class	CGR Based Features			FCGR Based Features			Hybrid (CGR + FCGR) Features		
	P _r	Recall	F _s	P _r	Recall	F _s	P _r	Recall	F _s
Normal	0.93	0.95	0.94	0.89	0.96	0.92	0.91	0.97	0.94
Osteopenia	0.92	0.90	0.91	0.93	0.87	0.90	0.95	0.91	0.93
Osteoporosis	0.92	0.90	0.91	0.89	0.86	0.87	0.93	0.87	0.90
Macro Average	0.92	0.92	0.92	0.90	0.89	0.90	0.93	0.92	0.92
Weighted Average	0.92	0.92	0.92	0.91	0.90	0.90	0.93	0.93	0.93

blocks, Bidirectional GRU (BiGRU) layers, and an attention mechanism with a context vector. Reduced variants were constructed by excluding one component at a time to analyze its impact on classification performance.

All experiments were conducted using an 80:20 patient-wise train-test split, and performance was assessed using accuracy, precision, recall, and F1-score. The evaluation was carried out separately for binary and multiclass classification tasks, with results summarized in Tables 8 and 9.

For binary classification, the complete model achieves the highest performance across all feature representations. Using hybrid features, the full model attains an accuracy of 0.98, whereas removing the Transformer module reduces the accuracy to approximately 0.94 - 0.95, corresponding to a drop of about 3 - 4%. Similarly, excluding the BiGRU layer results in a reduction of nearly 2 - 3%, indicating the importance of sequential feature modeling. The removal of the attention mechanism leads to a performance decrease of around 2%, demonstrating its role in emphasizing discriminative feature regions. In contrast, the baseline model (Conv1D + Dense) shows a substantial decline, with accuracy dropping to approximately 0.83, reflecting a degradation of nearly 15% compared to the full model. These observations confirm that each component contributes to improving classification performance.

The multiclass classification problem shows a similar pattern, with the whole hybrid model incorporating fused features achieving an accuracy of 0.92. When the Transformer component is removed, the accuracy decreases to approximately 0.88 - 0.89, corresponding to a drop of about 3 - 4%. Likewise, removing the BiGRU layer

Table 8: Ablation Study Results for Binary Class Classification

Feature	Model A (No Transformer) [Conv1D + BiGRU + Attention]				Model B (No BiGRU) [Conv1D + Transformer + Attention]				Model C (No Attention) [Conv1D + Transformer + BiGRU]				Model D (Conv + Dense) [Conv1D + Dense]				Model E (Conv + Transformer) [Conv1D + Transformer]				HTGCA-Net (Full Model)			
	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s
CGR	0.91	0.89	0.89	0.88	0.93	0.94	0.93	0.93	0.92	0.92	0.91	0.91	0.80	0.81	0.80	0.80	0.86	0.87	0.87	0.87	0.96	0.98	0.97	0.97
FCGR	0.92	0.92	0.90	0.90	0.94	0.93	0.94	0.93	0.93	0.92	0.92	0.93	0.81	0.81	0.82	0.80	0.88	0.87	0.88	0.88	0.97	0.97	0.97	0.98
Hybrid	0.94	0.92	0.92	0.93	0.96	0.95	0.94	0.95	0.96	0.96	0.97	0.96	0.83	0.84	0.83	0.82	0.90	0.89	0.90	0.91	0.98	0.98	0.99	0.98

Table 9: Ablation Study Results for Multiclass Classification

Feature	Model A (No Transformer) [Conv1D + BiGRU + Attention]				Model B (No BiGRU) [Conv1D + Transformer + Attention]				Model C (No Attention) [Conv1D + Transformer + BiGRU]				Model D (Conv + Dense) [Conv1D + Dense]				Model E (Conv + Transformer) [Conv1D + Transformer]				HTGCA-Net (Full Model)			
	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s	A _c	P _r	R _c	F _s
CGR	0.87	0.88	0.87	0.87	0.88	0.89	0.87	0.88	0.89	0.90	0.88	0.89	0.81	0.81	0.80	0.80	0.85	0.85	0.84	0.84	0.91	0.92	0.91	0.92
FCGR	0.86	0.87	0.85	0.86	0.87	0.88	0.86	0.87	0.88	0.89	0.87	0.88	0.80	0.79	0.79	0.80	0.84	0.85	0.83	0.84	0.90	0.90	0.89	0.89
Hybrid	0.88	0.89	0.87	0.88	0.89	0.90	0.88	0.89	0.90	0.91	0.90	0.89	0.83	0.84	0.83	0.83	0.86	0.87	0.85	0.85	0.92	0.93	0.91	0.92

results in a reduction of around 2 - 3%, while excluding the attention mechanism leads to a decrease of approximately 2%. The baseline model again exhibits the lowest performance, with accuracy values around 0.83, indicating a degradation of nearly 9 - 10% compared to the full architecture. Notably, the impact of component removal is more pronounced in the multiclass scenario, as distinguishing between Normal, Osteopenia, and Osteoporosis requires more discriminative and context-aware feature representations.

Overall, the ablation analysis demonstrates that the integration of convolutional feature extraction, Transformer-based global modeling, sequential learning via BiGRU, and attention-based feature refinement leads to consistent and measurable performance gains. The combined architecture provides a balanced and robust framework for both binary class and multiclass osteoporosis classification.

These quantitative results further confirm that each architectural component contributes incrementally to the overall performance, with the hybrid configuration yielding the most consistent improvements across different feature representations.

7 Comparative Analysis

Tables 6 and 7, which summarize the experimental results, show that both CGR-based and FCGR-based feature representations perform competitively in binary class(Normal vs. Osteoporosis) and multiclass (Normal, Osteopenia, Osteoporosis) classification tasks. CGR and FCGR feature sets produced similar test accuracies of 0.97 for the 80:20 data split in binary classification, but the hybrid representation achieved a slightly higher accuracy of 0.98, indicating its capacity to better capture discriminative patterns.

Figure 26 provides a comparison of performance metrics, such as Accuracy, Precision, Recall, and F1-score. The hybrid feature representation consistently shows

stable and balanced performance across all evaluation parameters. Figure 27 provides additional comparative analysis against existing methods, evaluating the proposed approach alongside previous researches like Valentinitich *et al.* [17], Sela *et al.* [44], and Jang *et al.* [45]. The results indicate that the proposed method achieves competitive accuracy (0.98) while maintaining a favorable balance between sensitivity and specificity, which is essential for reliable clinical decision support. In addition, Figure 28 provides a comparison with several established deep learning architectures, including MobileNetV2, ResNet-50, and EfficientNet. The proposed method demonstrates consistent improvements in classification performance, which can be attributed to its ability to jointly capture spatial structural characteristics and frequency-based statistical information.

Figures 8 to 15 present the confusion matrices, ROC curves, and accuracy-loss plots for different feature representations. Specifically, Figures 8, 10, and 9 correspond to CGR-based features, Figures 11, 13, and 12 represent FCGR-based features, and Figures 14, 16, and 15 illustrate the hybrid feature approach. For the 80:20 and 90:10 dataset distribution, the ROC analysis shows consistently high AUC values across all feature representations, demonstrating the suggested model’s good discriminative capacity. The stability of AUC across different dataset distributions further suggests that the model maintains reliable performance under varying training-testing configurations.

Additionally, the comparative analysis presented in Figure 27 demonstrates that the proposed approach consistently improves several evaluation metrics. In particular, sensitivity and specificity exhibit notable gains, contributing to a more balanced classification outcome. Similarly, Figure 28 highlights improvements in Precision, Recall, and F1-score compared with several state-of-the-art models. All comparisons were conducted under consistent evaluation settings where applicable. Together, these findings indicate that the proposed feature representation provides an effective and reliable framework for osteoporosis classification from X-ray images.

Similarly, Table 7 summarizes the performance of the suggested feature representations for multiclass classification for the 80:20 dataset split. The relative efficacy of CGR-based, FCGR-based, and hybrid (CGR + FCGR) feature sets was assessed. The results show that both CGR and FCGR feature representations achieve comparable performance, yielding test accuracies of approximately 0.91 and 0.90, respectively. In comparison, the hybrid feature representation attains a slightly higher accuracy of 0.92. This improvement suggests that the integration of CGR and FCGR features provides a more comprehensive characterization of the data, leading to enhanced classification performance.

Although individual feature sets capture important characteristics of trabecular bone structure, their standalone application does not consistently provide balanced performance across all evaluation metrics and data splits. In contrast, the hybrid feature representation, which integrates spatial fractal patterns from CGR with frequency-based statistical information from FCGR, offers a more comprehensive characterization of bone texture. This complementary integration enables the model to better capture both local structural variations and global distribution patterns within the X-ray images. As a result, the hybrid approach demonstrates more stable and

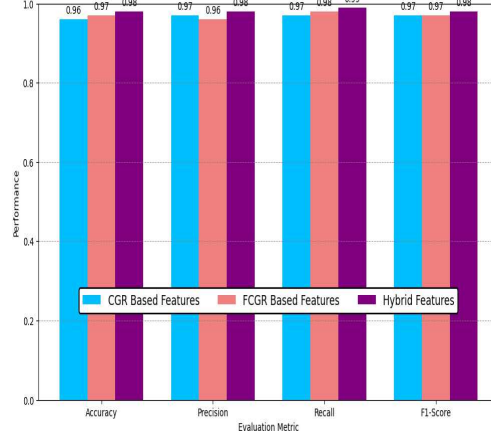


Fig. 26: Evaluation of Accuracy, Precision, Recall, and F1-Score in comparison with CGR Based Features, FCGR Based Features and Hybrid Features for Binary Class Classification

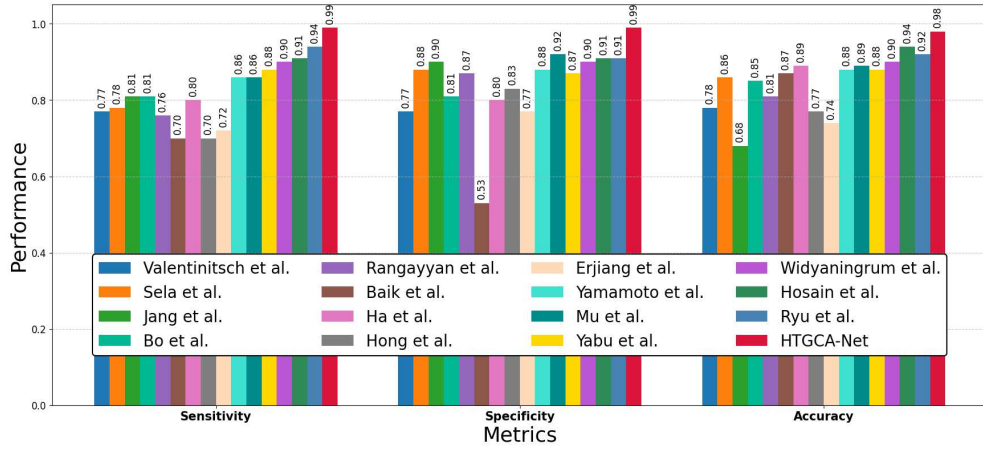


Fig. 27: Performance Comparison in terms of Sensitivity, Specificity and Accuracy with proposed model and various researches for Binary-Class Classification

consistent performance across different split configurations, with improved classification accuracy and a more balanced precision recall trade-off. In particular, the hybrid representation is more effective in addressing the challenging distinction between Osteopenia and Osteoporosis classes, where subtle structural differences are present.

Overall, these findings demonstrate that integrating CGR and FCGR features improves the model's discriminative capability by providing a richer feature representation, thereby enhancing generalization across different data distributions. The

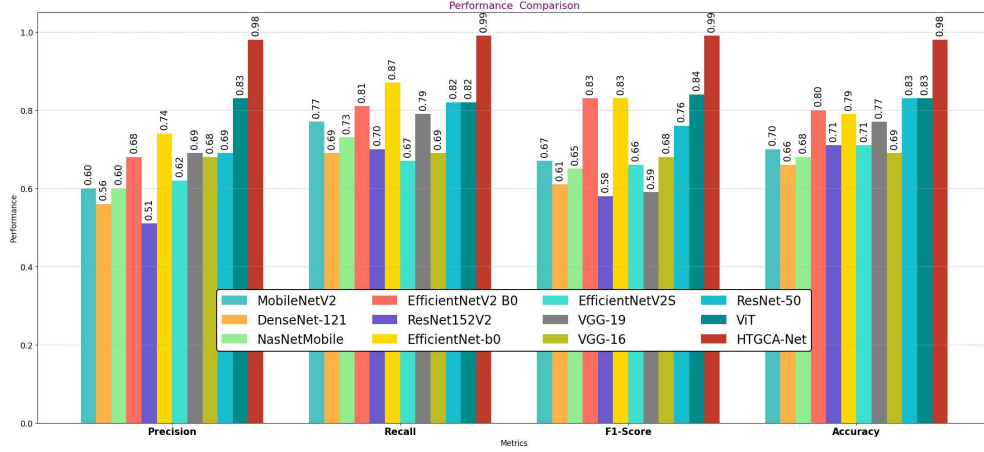


Fig. 28: Performance Comparison in terms of Precision, Recall, F1-Score and Accuracy with various State-Of-The-Art Models for Binary-Class Classification

comparative analysis in Figure 29, which shows competitive performance across sensitivity, specificity, and accuracy metrics, further demonstrates the effectiveness of the proposed approach. Similarly, Figure 31 highlights improvements in Precision, Recall, and F1-score compared with several state-of-the-art models. All comparisons were conducted under consistent evaluation settings where applicable. Additionally, Figures 17 to 25 present the confusion matrices, loss-accuracy graphs and ROC curves for the multiclass classification task. Specifically, Figures 17, 18 and 19 correspond to CGR-based features, Figures 20, 21 and 22 represent FCGR-based features, and Figures 23, 24 and 25 illustrate the hybrid feature representation. The ROC analysis indicates that the hybrid approach achieves higher AUC values and improved class separability, confirming its robustness in distinguishing between Normal, Osteopenia, and Osteoporosis categories.

The comparative analysis presented in Figure 27 indicates that the proposed HTGCA-Net achieves consistent improvements over existing methods in binary classification, with relative gains of 4.26%-44.12% in accuracy, 5.32%-41.43% in sensitivity, and 7.61%-86.79% in specificity. Similarly, as illustrated in Figure 28, the proposed model demonstrates notable performance enhancements compared to state-of-the-art approaches, with improvements ranging from 18.07%-92.16% in precision, 13.79%-47.76% in recall, 17.86%-70.69% in F1-score, and 18.07%-50.77% in accuracy. For the multiclass classification task, the results in Figure 29 show that HTGCA-Net achieves consistent gains across evaluation metrics, with improvements of 2.22%-46.03% in sensitivity, 1.09%-47.62% in specificity, and 2.22%-46.03% in accuracy compared to existing methods. Furthermore, Figure 31 highlights that the proposed model outperforms recent state-of-the-art techniques, achieving improvements of 22.37%-82.35% in precision, 5.75%-37.31% in recall, 10.84%-58.62% in F1-score, and 10.84%-39.39% in accuracy. These results collectively demonstrate the effectiveness of the proposed

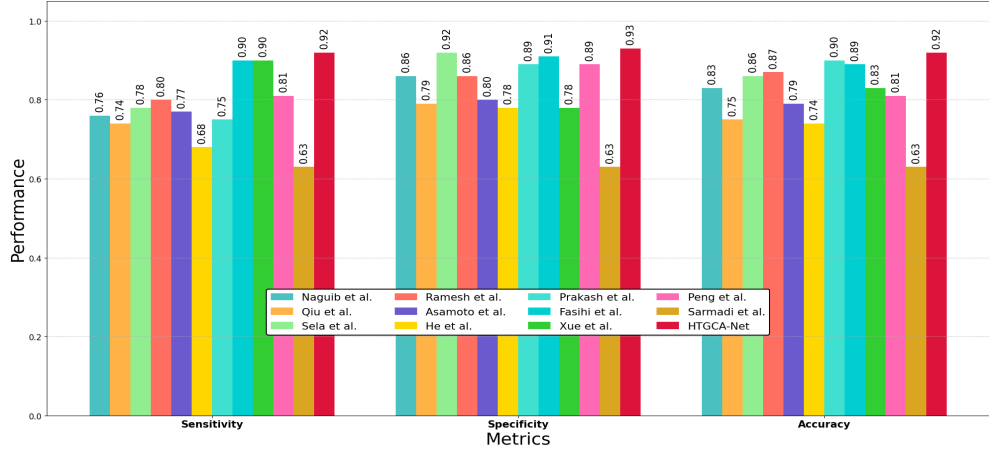


Fig. 29: Performance Comparison in terms of Sensitivity, Specificity and Accuracy with proposed model and various researches for Multi-Class Classification

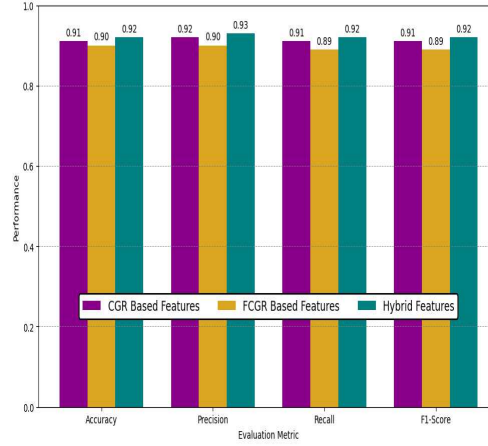


Fig. 30: Evaluation of Accuracy, Precision, Recall, and F1-Score in comparison with CGR Based Features, FCGR Based Features and Hybrid Features for Multi-Class Classification

framework in achieving robust and reliable performance across both binary class and multiclass classification scenarios.

8 Conclusion

This research presents a robust classification framework for osteoporosis detection using plain radiographic images of the hip regions, incorporating both publicly available and clinically acquired datasets. A hybrid feature extraction strategy is introduced

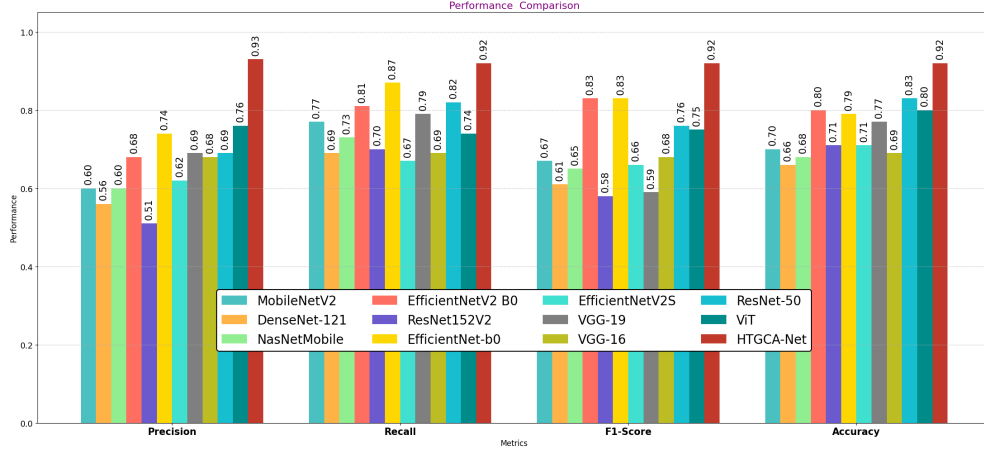


Fig. 31: Performance Comparison in terms of Precision, Recall, F1-Score and Accuracy with various State-Of-The-Art Models for Multi-Class Classification

by integrating Chaos Game Representation (CGR) and Frequency Chaos Game Representation (FCGR), enabling the capture of complementary structural and statistical characteristics of trabecular bone. The CGR-based features effectively model spatial and fractal patterns derived from symbolic encoding of grayscale intensities, while FCGR-based features quantify the frequency distribution of these patterns, providing a global representation of bone texture.

For classification, a hybrid deep learning architecture is employed, combining convolutional feature extraction, Transformer encoder blocks, sequential modeling through Bidirectional GRU layers, and an attention mechanism with context vector. This integrated architecture improves feature discrimination, reduces the impact of irrelevant patterns, and makes it easier to learn both local and global dependencies. By highlighting structural changes that are clinically significant, the inclusion of attention processes enhances interpretability even further.

The suggested framework achieves consistent and balanced performance across assessment measures, including as accuracy, sensitivity, specificity, and F1-score, for both binary class and multiclass classification tasks, according to extensive experiments carried out under various dataset split configurations. The results indicate that the suggested deep learning architecture in conjunction with the hybrid CGR-FCGR feature representation offers a dependable and computationally effective method for osteoporosis detection. Overall, the suggested system offers an effective approach for precise and scalable osteoporosis screening based on widely available X-ray imaging, demonstrating the efficacy of combining symbolic feature modeling with attention-driven deep learning.

9 Limitation and Future Work

The proposed hybrid CGR-FCGR framework demonstrates promising performance for osteoporosis classification, certain limitations should be acknowledged. First, the current research focuses on discrete class prediction (Normal, Osteopenia, and Osteoporosis) rather than direct estimation of Bone Mineral Density (BMD). Since clinical diagnosis of osteoporosis is fundamentally based on quantitative BMD measurements, the proposed approach is intended as a supportive screening tool rather than a substitute for standard diagnostic procedures.

Second, the feature representation is derived from symbolic encoding of grayscale X-ray images, which may not fully capture all anatomical and physiological variations present in complex bone structures. Additionally, while the use of a hybrid deep learning architecture improves feature discrimination, it introduces a moderate level of model complexity that may require careful optimization for deployment in resource-constrained environments.

Future work will focus on extending the proposed framework toward regression-based modeling for direct BMD estimation, enabling a more comprehensive and clinically relevant assessment. Furthermore, the integration of clinically interpretable visualization techniques will be explored to enhance transparency and facilitate adoption in real-world settings. Validation on larger, multi-center datasets with diverse demographic distributions will also be conducted to further evaluate the generalization capability and robustness of the proposed method.

Dataset Availability Statement

The datasets generated and/or analysed during the current research are not publicly available due to regulatory, ethical, and data-sharing restrictions associated with an ongoing Indian Council of Medical Research (ICMR) Intermediate Extramural Research Project Grant (Project No. IIRPIG-2024-01-00483). The dataset comprises clinical hip radiographic images collected from multiple participating institutions under approved research protocols. The datasets will be considered for public release upon completion of the project period in 2027, subject to institutional, ethical, and regulatory approvals. Until then, the datasets are available from the corresponding author on reasonable request and with permission from the relevant authorities.

Acknowledgment

The authors gratefully acknowledge the financial support provided by the Indian Council of Medical Research (ICMR) under the Intermediate Extramural Project Grant (Project No. IIRPIG-2024-01-00483), titled "Development of Low-Cost Artificial Intelligent System for Prediction of Bone Mineral Density (BMD) From Plain Radiographic Images in Osteoporosis Detection". This support has been instrumental in carrying out the present research work.

Funding Declaration

No direct funding was received for the present study. The dataset used in this research was collected under the Indian Council of Medical Research (ICMR) Intermediate Extramural Project Grant (Project No. IIRPIG-2024-01-00483), titled "Development of Low-Cost Artificial Intelligent System for Prediction of Bone Mineral Density (BMD) From Plain Radiographic Images in Osteoporosis Detection".

Clinical Trial

Not applicable. This study did not involve a clinical trial.

Consent to Participate declaration

Not applicable

Consent to Publish declaration

Not applicable.

Ethics declaration

The dataset used in this research was collected under the Indian Council of Medical Research (ICMR) Intermediate Extramural Project Grant (Project No. IIRPIG-2024-01-00483), titled "Development of Low-Cost Artificial Intelligent System for Prediction of Bone Mineral Density (BMD) From Plain Radiographic Images in Osteoporosis Detection." Image acquisition and data collection were conducted following an approved clinical protocol within the scope of the ICMR-funded project.

References

- [1] J. A. Kanis, "Assessment of osteoporosis at the primary health-care level," *World Health Organization Collaborating Centre*, 2008.
- [2] O. Johnell and J. A. Kanis, "An estimate of the worldwide prevalence and disability associated with osteoporotic fractures," *Osteoporosis International*, vol. 17, no. 12, pp. 1726–1733, 2006.
- [3] S. Baim et al., "Official positions of the International Society for Clinical Densitometry," *Journal of Clinical Densitometry*, vol. 11, no. 1, pp. 75–91, 2008.
- [4] M. Singh et al., "Changes in trabecular pattern of the upper end of the femur as an index of osteoporosis," *Journal of Bone and Joint Surgery*, vol. 52, no. 3, pp. 457–467, 1970.
- [5] H. W. Jang et al., "Deep learning-based osteoporosis classification using hip radiographs," *IEEE Access*, vol. 9, pp. 123456–123467, 2021.

- [6] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” *International Conference on Learning Representations (ICLR)*, 2021.
- [7] K. He et al., “Deep residual learning for image recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [8] A. Jennane et al., “Fractal dimension and lacunarity analysis of trabecular bone in osteoporosis,” *IEEE Transactions on Medical Imaging*, vol. 20, no. 10, pp. 1024–1030, 2001.
- [9] S. Majumdar et al., “Fractal analysis of bone structure: A review,” *Medical Physics*, vol. 26, no. 9, pp. 1940–1950, 1999.
- [10] H. J. Jeffrey, “Chaos game representation of gene structure,” *Nucleic Acids Research*, vol. 18, no. 8, pp. 2163–2170, 1990.
- [11] F. Deschavanne et al., “Genomic signature: Characterization and classification of species assessed by chaos game representation,” *Molecular Biology and Evolution*, vol. 16, no. 10, pp. 1391–1399, 1999.
- [12] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [13] S. Arivazhagan and L. Ganesan, “Texture classification using wavelet transform,” *Pattern Recognition Letters*, vol. 24, no. 9–10, pp. 1513–1521, 2003.
- [14] R. Acharya et al., “Automated diagnosis of osteoporosis using texture features,” *Information Sciences*, vol. 178, no. 17, pp. 3275–3286, 2008.
- [15] H. F. Lochel and D. Heider, “Chaos game representation and its applications in bioinformatics,” *Computational and Structural Biotechnology Journal*, vol. 19, pp. 6263–6271, 2021.
- [16] H. J. Jeffrey, “Chaos game representation of gene structure,” *Nucleic Acids Research*, vol. 18, no. 8, pp. 2163–2170, 1990.
- [17] A. Valentinitich et al., “Opportunistic osteoporosis screening in multi-detector CT images via local classification of textures,” *Osteoporosis International*, vol. 30, pp. 1275–1285, 2019.
- [18] S. Bhattacharya et al., “Computer based automatic detection and classification of osteoporosis in bone radiographs,” in *Proc. 6th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, 2019.
- [19] J.-G. Shim et al., “Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women,” *Archives of Osteoporosis*, vol. 15, pp. 1–9, 2020.

- [20] S. Mu, J. Wang, and S. Gong, "Application of medical imaging based on deep learning in the treatment of lumbar degenerative diseases and osteoporosis with bone cement screws," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 1, p. 2638495, 2021.
- [21] A. Yabu *et al.*, "Using artificial intelligence to diagnose fresh osteoporotic vertebral fractures on magnetic resonance images," *The Spine Journal*, vol. 21, no. 10, pp. 1652–1658, 2021.
- [22] R. Dzierzak and Z. Omiotek, "Application of deep convolutional neural networks in the diagnosis of osteoporosis," *Sensors*, vol. 22, no. 21, p. 8189, 2022.
- [23] L. Liu *et al.*, "A hierarchical opportunistic screening model for osteoporosis using machine learning applied to clinical data and CT images," *BMC Bioinformatics*, vol. 23, no. 1, p. 63, 2022.
- [24] G. A. Albuquerque *et al.*, "Osteoporosis screening using machine learning and electromagnetic waves," *Scientific Reports*, vol. 13, no. 1, p. 12865, 2023.
- [25] M. J. A. Rasool *et al.*, "KONet: Toward a weighted ensemble learning model for knee osteoporosis classification," *IEEE Access*, vol. 12, pp. 5731–5742, 2024.
- [26] B. Suh, S. Y. Jung, Y. Kim, S. M. Kwon, Y. Lee, and J. H. Lee, "Interpretable deep-learning approaches for osteoporosis risk screening and individualized feature analysis using large population-based data: Model development and performance evaluation," *Journal of Medical Internet Research*, vol. 25, p. e40179, 2023.
- [27] M. H. Belali, M. S. Hossain, M. Alazab, H. Alsulami, and F. Ali, "Bone osteoporosis fractures detection with deep learning: An X-ray image analysis approach," *Biomedical Materials & Devices*, pp. 1–19, 2025.
- [28] D. Srinivasan, A. Ramesh, S. Prakash, M. P. Kumar, and R. Kumar, "Bonevoyage: Navigating the depths of osteoporosis detection with a dual-core ensemble of cascaded ShuffleNet and neural networks," *Journal of X-Ray Science and Technology*, vol. 33, no. 1, pp. 3–25, 2025.
- [29] L. He *et al.*, "Radiomics Based on Lumbar Spine Magnetic Resonance Imaging to Detect Osteoporosis," *Academic Radiology*, vol. 28, no. 6, pp. e165–e171, 2021.
- [30] U. M. Prakash, M. C. Durai, and M. A. Durai, "4x-expert systems for early prediction of osteoporosis using multi-model algorithms," *Measurement*, vol. 180, p. 109543, 2021.
- [31] Y. Fang *et al.*, "Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks," *European Radiology*, vol. 31, pp. 1831–1842, 2021.

- [32] Q. Dong *et al.*, “Deep learning classification of spinal osteoporotic compression fractures on radiographs using an adaptation of the genant semiquantitative criteria,” *Academic Radiology*, vol. 29, no. 12, pp. 1819–1832, 2022.
- [33] L. Fasihi, D. Asemani, A. M. Jaladat, and A. Jafari, “Artificial intelligence used to diagnose osteoporosis from risk factors in clinical data and proposing sports protocols,” *Scientific Reports*, vol. 12, no. 1, p. 18330, 2022.
- [34] L. Xue *et al.*, “A dual-selective channel attention network for osteoporosis prediction in computed tomography images of lumbar spine,” *Acadlore Transactions on AI and Machine Learning*, vol. 1, no. 1, pp. 30–39, 2022.
- [35] K. Zhang *et al.*, “End to end multitask joint learning model for osteoporosis classification in CT images,” *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 3018320, 2023.
- [36] Y.-C. Chen *et al.*, “Automatic segmentation and radiomic texture analysis for osteoporosis screening using chest low-dose computed tomography,” *European Radiology*, vol. 33, no. 7, pp. 5097–5106, 2023.
- [37] M. U. Ozic, M. Tassoker, and F. Yuce, “Fully automated detection of osteoporosis stage on panoramic radiographs using YOLOv5 deep learning model and designing a graphical user interface,” *Journal of Medical and Biological Engineering*, vol. 43, no. 6, pp. 715–731, 2023.
- [38] T. Peng *et al.*, “A study on whether deep learning models based on CT images for bone density classification and prediction can be used for opportunistic osteoporosis screening,” *Osteoporosis International*, vol. 35, no. 1, pp. 117–128, 2024.
- [39] A. Sarmadi *et al.*, “Comparative analysis of vision transformers and convolutional neural networks in osteoporosis detection from X-ray images,” *Scientific Reports*, vol. 14, no. 1, p. 18007, 2024.
- [40] Y. Wu *et al.*, “Artificial intelligence assisted automatic screening of opportunistic osteoporosis in computed tomography images from different scanners,” *European Radiology*, vol. 35, no. 4, pp. 2287–2295, 2025.
- [41] B. Zhang, D. Tian, Z. Yu, D. Zeng, Y. Zhou, Y. Lu, and Y. Zhang, “Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study,” *Bone*, vol. 140, p. 115561, 2020.
- [42] J. Pan, X. Zhang, Y. Zhou, Y. Liu, H. Sun, H. Yang, Y. Gao, Y. Huang, and Z. Wang, “Feasibility study of opportunistic osteoporosis screening on chest CT using a multi-feature fusion DCNN model,” *Archives of Osteoporosis*, vol. 19, no. 1, pp. 1–13, 2024.

- [43] L. Dinda, J. Pradhan, T. K. Dey, and M. Raj, "Fusion of ViT-Based Deep Features and Phi-Grid Handcrafted Features for Osteoporosis Detection Using Attention-Infused ResNet MLP and Autoencoder," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3634211.
- [44] E. I. Sela and R. Pulungan, "Osteoporosis identification based on the validated trabecular area on digital dental radiographic images," *Procedia Computer Science*, vol. 157, pp. 282–289, 2019.
- [45] R. Jang *et al.*, "Prediction of osteoporosis from simple hip radiography using deep learning algorithm," *Scientific Reports*, vol. 11, no. 1, p. 19997, 2021.