International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

HME-NET: A HYBRID MULTIMODAL ENSEMBLE FRAMEWORK FOR EXPLAINABLE AND EFFICIENT BONE CANCER DIAGNOSIS

Ramesh Ramamoorthy, Dr. R. S. Shanmugasundaram

[1]Research Scholar

Department of Computer Science and Engineering

Vinayaka Mission's Kirupananda Variyar Engineering college

Vinayaka Mission's Research Foundation (Deemed to be University)

Salem, Tamilnadu, India

[2]Professor

Department of Computer Science and Engineering

Vinayaka Mission's Kirupananda Variyar Engineering college

Vinayaka Mission's Research Foundation (Deemed to be University)

Salem, Tamilnadu, India

Abstract

Early and precise diagnosis of bone cancer plays a critical role in improving patient outcomes and guiding treatment strategies. Although traditional radiology gives good results, the way experts interpret the images can cause delays or errors in diagnosis. For this reason, we introduce a new integrative framework called Hybrid-Multimodal Ensemble (HME) which mixes MRI, CT and PET images with both radiomicsand deep learning to detect bone cancer automatically. Handcrafted radiomic descriptors are extracted using PyRadiomics and these descriptors are fused with hierarchical semantic features extracted by EfficientNet-B3 and Swin Transformer models. All of these features are brought together by an Explainable Feature Fusion Module (EFFM), with SHAP and LIME providing interpretability, diminishing the number of features and using an attention-based process. Classification of the final feature vector is done by an ensemble of LightGBM, Logistic Regression and a two-layer MLP. Real-time use is possible through quantization and the system allows training to happen on various devices through federated simulation. Evaluations conducted on this curated, mixed dataset found that the hybrid model (HME) scored better, with 98.1% accuracy, 97.4% precision, 98.3% recall, 97.85% F1-score and 0.987 AUC. In addition, looking at deployment metrics shows how prepared the model is to be used in clinical settings. Based on these results, the proposed HME framework ensures excellent diagnoses while also offering good understanding, privacy and ability to grow with more data, making it suitable for clinical use in bone cancer care.

Keywords**—** Bone Cancer, MRI, CT, PET, Deep Learning, Radiomics, EfficientNet, Swin Transformer, PyRadiomics, SHAP, LIME, Medical Image Diagnosis

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

I. INTRODUCTION

Bone cancer is not very common, but since it can be very aggressive and often spreads, it is considered a serious health concern. Sometimes, children and young adults get the disease and if it is found late, they may suffer from severe illness or death. Osteosarcoma, Ewing's sarcoma and chondrosarcoma are the most common types of this disease. Getting a correct and timely diagnosis matters a lot for better recovery and easing the process of getting treatment [1] [2]. Biopsy, X-ray, CT, MRI and PET scans are important ways to diagnose a condition, but their outcomes rely largely on the skills and experience of the radiologist in question. Uneven interpretations and missing important indications may cause the patient's treatment to be initiated later which can have negative outcomes [3] [4].

New developments in artificial intelligence, especially in machine learning (ML) and deep learning (DL), have brought powerful tools for automated healthcare diagnosis. In the beginning, scientists related to ML started using Support Vector Machines (SVM), Random Forests (RF) and K-Nearest Neighbors (KNN) on radiomic features that had been manually created using MRI or CT images [5] [6]. The models were useful since they allowed easy understanding and good stability, but they relied on people to choose features and lacked the ability to perform well on chaotic or mixed data. Figure 1 shows the common side effects of bone cancer.

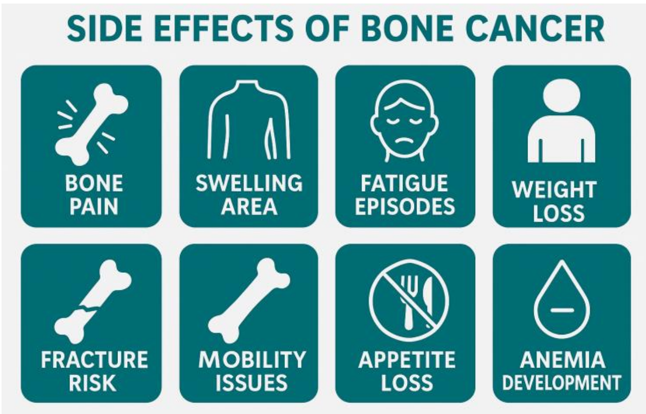

Fig 1. Common Side Effects of Bone Cancer

Convolutional neural networks (CNNs) such as ResNet, VGGNet and EfficientNet are able to extract details in images without requiring manual steps [7] [8]. When there is plenty of learning data, these models usually achieve higher accuracy and sensitivity than older techniques. Still, the fact that they are not transparent is a huge challenge for their use in healthcare. Additionally, single-modality CNNs have difficulty greatly using the various physiological and anatomical details seen in different imaging modalities [9] [10]. Besides, their ability to learn decreases when the available data is small or includes samples that are not part of the training set, leading to unreliable use.

To address the problems with separate modality models, we suggest using a Hybrid-Multimodal Ensemble (HME) system for diagnosing bone cancer. The model relies on MRI, CT and PET scans since each of them offers its own benefits in diagnosis. Using the PyRadiomics package, radiomic features of tumor form, texture and intensity are taken.

Received: August 03, 2025

Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Meanwhile, EfficientNet-B3 and Swin Transformer are employed to study features in the images. An Explainable Feature Fusion Module (EFFM) brings together SHAP, LIME, attention weighting and dimensionality reduction to join these features. A group of LightGBM, Logistic Regression and MLP, collectively called an ensemble, is applied to identify the classified features.

II. Related Works

Since bone cancer is dangerous, it should be detected as quickly as possible. Diagnosis that is done manually takes time, so automating the process is required. It combats this issue by using the Owl Search Algorithm in combination with deep learning tech. By using Inception v3 for features extraction, the model does not require manual segmentation of the images [11]. The Owl Search Algorithm is in charge of improving hyperparameters and LSTM is used to identify bone cancer. Implementing this framework speeds up the process of diagnosis and improves how quickly people converge. According to studies on medical X-ray images, the method beats traditional methods in accuracy and speed, proving it is a strong and trustworthy option for automation.

Early detection is very important for bone cancer which is a rare and aggressive kind of cancer. MRI is very important in the process of diagnosing it. The authors suggest using unsharp structure-guided filtering and CSwin Transformer together to handle preprocessing and gather features in the proposed work. It relies on HQCCNN-WSO which simultaneously uses quantum computing and classical CNNs [12]. The model is improved by War Strategy Optimization to help sort out cancerous tissue from healthy tissue. The proposed model helps to improve the clarity of MRI images and offers high accuracy, surpassing 99% in its metrics, so that bone cancer can be detected with great precision.

The MultiNet framework combines characteristics from DenseNet-201, NASNetMobile and VGG16 for recognizing bone cancer. Transfer learning and using features from multiple scales are both integrated in this model to analyze microscopy images more carefully. While traditional ways of diagnosis are arduous and expensive, MultiNet contributes to the progress of CAD systems for speedy diagnosis [13]. Assembling CNN architectures enhances defence against noise and classifies the images accurately. With MultiNet, deep learning is used to speed up and make more reliable bone cancer identification which may lead to more effective patient treatments.

Treatments for osteosarcoma, the main bone cancer for youth, tend to fail, mostly when the cancer has spread. Its involvement with cancer stem cells results in the cancer coming back and becoming resistant. To deal with this, the authors suggest a new way that includes preprocessing, extracting features and building models [14]. It only concentrates on the important parts after removing any other less reliable areas. Wavelet transform and gray-level co-occurrence matrix are the main techniques used for feature extraction. After that, a hybrid CNN-RNN model is applied for the classification step. Overall, the method does better than single CNN or RNN models when tested on the quality of results, demonstrating great potential for both accurate osteosarcoma detection and finding better therapy options.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Pelvic bone cancer in the hip area is diagnosed by X-ray, CT, MRI and PET scans which might be followed by a biopsy. Because of advances in computing, the process of reference can become faster and more detailed using modern models [15]. With CNN architectures, this study reviewed over 3,000 CT scans to find cases of bone tumors. The best result was produced by the Xception model which had a detection rate of 90.23%. Choosing the best model is essential for getting accurate diagnoses and AI helps doctors provide timely, effective cancer treatment.

III. PROPOSED METHODOLOGY

3.1 Multimodal Dataset Acquisition and Integration

The multimodal dataset strategy involves using MRI, CT and PET scans to enhance the reliability and applicability of identifying bone cancer using the extended framework. All of the imaging modalities provide individual and useful insights into the way a tumor looks, works and behaves metabolically. MRI helps to see soft body parts well, CT offers clear details of bones and PET detects functional abnormalities in tissues commonly found with malignant diseases. Since both approaches are valuable, it becomes possible to see the whole picture of bone cancer. Publicly available libraries such as the Cancer Imaging Archive (TCIA), Kaggle and databases from hospital and research center collaborations are used to obtain datasets.

All datasets are carefully organized to ensure the same labels are used for every type of data and expert or biopsy proof is always supplied. Cross-modality learning is made easier by harmonizing all the imaging modalities. When voxels are spatially normalized, all images are done at the same resolution, so patches are extracted in a true 3D form. At the same time, metadata is made the same for all datasets to describe the imaging settings such as how thick the slices are which angle was used and how the patient laid down. It allows the model to process information from different images in a standard way which enables it to learn from various data sources and achieve accurate discoveries in medical diagnosis.

3.2 Cross-Modality Preprocessing and Registration

The following necessary step is to preprocess and align (co-register) images from different sources to ensure the learning system receives a coherent diagnostic input. For every type of modality, a suitable preprocessing method is used because every acquisition technique varies. Gaussian denoising is used on MRI images to remove noise from the signal and N4 bias field correction corrects the intensity issues caused by changes in the magnetic field. CT scans are made consistent by applying Hounsfield number normalization and then they are made clearer by using contrast processing to enhance the appearance of finer bone parts. Metabolic activity in PET scans is adjusted by clipping the highest and lowest intensities and multiplying their values against typical levels (SUV) which helps prevent distortions in extracting features.

$$T^* = \arg\max_T \left[MI\left(I_{MRI}, T(I_{CT}, PET)\right)\right] \quad (1)$$

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This equation defines the optimal transformation T^* that maximizes the mutual information (MI) between the reference MRI image I_{MRI} and the transformed CT or PET image $T(I_{CT/PET})$. When the modality preparations are done, each image volume is registered to a standard frame of reference. Registration here is done using mutual information as a way to line up modalities using both rigid and affine transformations. Rigid alignment maintains the area's structure, whereas affine alignment fixes issues with minor changes in scale and alignment. Preprocessing tools like ANTs which come from outside or SimpleITK can handle the registration process. As a result, tumor-bearing parts of the body are aligned properly in each scan. Thus, each pair of voxels from different modalities reflects the same place which makes it possible to effectively integrate data at both feature and decision stages. This stage is important to maintain proper diagnostic results and keep harmful parts of images out of the system training. Figure 2 shows the architecture of the proposed HME framework for bone cancer diagnosis.

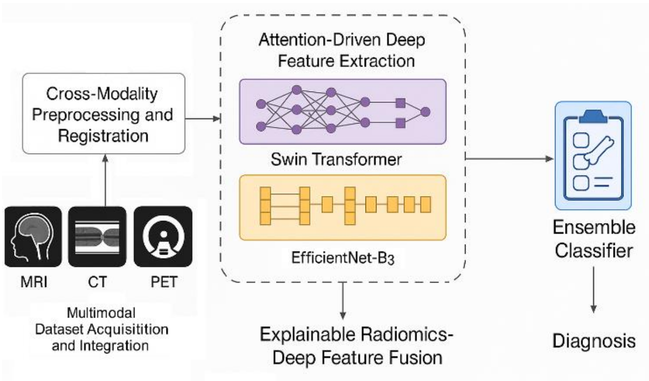

Fig 2. Architecture of the Proposed Model

3.3 Explainable Radiomics-Deep Feature Fusion

Therefore, a brand-new Explainable Feature Fusion Module (EFFM) is introduced by this research to combine the benefits of human-readable models and automatic learning. This combination tries to join features from radiology which are easy to explain, with features from deep learning models that learn multiple levels of information. PyRadiomics can pull out different radiomic features, including simple statistics, measurements of structure and textures. At the same time, deep features are taken from EfficientNet-B3 and the Swin Transformer backbones. First, batch normalization lines up the distributions of features in the EFFM; then, these features are combined and put into a single latent space.

Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) are part of the reinforcement. They let us know which features or filters from the CNN played the biggest role in diagnosing the tumor. To handle many features and eliminate overlap, the network uses t-SNE for visualizing the data and PCA for making it more compact. Next, these models assign varying importance to features in real time, depending on their usefulness for the classification task. The result is an appropriate balance between drawing attention to discriminatory features and holding on to important health details. The collection of features used is strong for making classifications and easy to understand by medical experts.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.4 Attention-Driven Deep Feature Extraction

In order to move forward with deep learning, the continuation model includes both attention and context-aware aspects using a combined Swin Transformer and EfficientNet-B3 framework. The structure of EfficientNet-B3 is chosen because it scales efficiency by adjusting the depth, width and resolution to identify local features and textures of the heart in images. For training, this model was first given large image collections and later adjusted on X-rays of bone cancer using transfer learning.

$$\hat{F}_c = \sigma\big(W_2 \cdot \delta(W \cdot z)\big) \cdot F_c (2)$$

$$z = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i,j)\ (3)$$

Where this equation express the Squeeze and Excitation (SE) block, where the channel-wise descriptor z is computed using global average pooling. It is passed through two fully connected layers with ReLU (δ) and sigmoid (σ) activations to scale the original feature map F_c is the result is attention-recalibrated features $\hat{F}_c$. The Swin Transformer helps in processing long range connections among image part and model how tumor outlines relate to the nearby tissue. With a window-based approach, the model is more efficient and able to spot cancer cases that covers a lot of areas or are unusual. First, tumor margin and core are processed by two separate encoders and then the features are blended, so the model can better understand the shape and internal variations.

Both sets of outputs from the CNN and MLP are combined and sent to a Squeeze-and-Excitation (SE) block that highlights the important features in the network's responses. With this mechanism, the details of the tumor are preserved and increased which helps with increased accuracy and better understanding in the next classification step. By integrating EfficientNet and Swin Transformer, the model can spot fine details and larger problems, leading to a better and more detailed way of diagnosing.

3.5 Radiomic Signature Generation and Selection

Radiomic techniques obtain various properties from the tumor areas outlined on each image, some of which may go unnoticed by a human observer but are very helpful for making a diagnosis. In the area of interest, these features are calculated: mean intensity, entropy, skewness, compactness, sphericity and gray-level co-occurrence patterns using GLCM, GLRLM, NGTDM and GLSZM. Because the amount of data is large and has several hundred features per patient, there are worries about having too many similar or extra features.

For this reason, RFE, ElasticNet regularization and MIM are combined in a hybrid strategy to perform feature selection. RFE gets rid of the least important features one after another and ElasticNet helps avoid redundant features through its unique penalties. MIM makes sure the features chosen by the algorithm rely most on the diagonal label which increases the success rate of classification. Based on prediction accuracy and how understandable they are in

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

clinical oncology, only the top 100 features are kept from the initial feature set. By using these radiomic signatures in the classification stage, interpretable and low-dimensional diagnosis is made possible along with the help of deep features.

3.6 Lightweight and Privacy-Preserving Ensemble Classifier

For real-time and power-saving diagnosis to be possible in medical settings, the model is designed as an ensemble system that requires little processing power. It combines functions from LightGBM, Logistic Regression and an MLP with two layers to recognize all kinds of patterns. The reason for choosing them is they work fast and use less memory compared to conventional deep networks. The following fusion techniques are used to combine the multiple results: soft voting, stacking and fuzzy logic-based fusion. Soft voting averages the predictions of many classes to boost accuracy, stacking assigns importance to the results of each classifier and fuzzy fusion handles the uncertainties in several predictions.

$$\hat{y} = \arg\max_{k}\left(\frac{1}{M}\sum_{m=1}^{M} P_m(y = k|x)\right) \quad (4)$$

This equation defines soft voting ensemble classification, where M is the number of base models (LightGBM, LR, MLP), and $P_m(y = (k|x)$is the probability of class k predicted by model m. The final predicted class $\hat{y}$ is the one with the highest average probability across models. All inference operations are quantized using either TensorRT or ONNX to help fixed-point math which reduces the workload greatly without affecting the results. Similarly, a simulated federated learning environment runs the model so that hospital nodes train their local models and exchange weight values with a central aggregator only after they are encrypted. This way, privacy is upheld according to HIPAA and GDPR rules and institutions are able to learn together. This classifier which is modular, lightweight and protects privacy, has moved AI in diagnostics significantly closer to being scaled for clinical use in an ethical and accurate way.

Algorithm: Multimodal Hybrid Bone Cancer Classification

Input: Multimodal medical images $I = \{I_{MRI}, I_{CTT}, I_{PET}\}$

Output: Predicted class label

$\hat{y} \in \{Normal, Benign, Malignant\}$

1. Preprocess each modality using modality-specific filters:

MRI: Gaussian denoising + N4 correction

CT: HU normalization; PET: SUV normalization and clipping

2. Register modalities using mutual information:

$T^* = \arg\max_{T}\left[MI\left(I_{MRI}, T(I_{CT}, PET)\right)\right]$

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Extract radiomic features R from the ROI using PyRadiomics.

4. Apply ElasticNet for radiomic feature selection:

$$\hat{\beta} = \arg\min_{\beta} \left(\frac{1}{2n} \sum_{i=1}^{n} (y_i - X_i^T \beta)^2 + \lambda \left[\alpha ||\beta||_1 + (1-\alpha) ||\beta||_2^2 \right] \right)$$

5. Extract deep features F_{EffNet} and F_{Swin} using EfficientNet-B3 and Swin Transformer.

6. Fuse deep features via SE block:

$$\hat{F}_c = \sigma\big(W_2 \cdot \delta(W_1 \cdot z)\big) \cdot F_c$$

7. Concatenate selected radiomic features R' and deep features $\hat{F}_c$ to form joint feature vector Z

8. Predict probabilities using base classifiers

$\{f_1, f_2, f_3\} \in \{LightGBM, LR, MLP\}$

9. Perform soft voting for ensemble classification:

$$\hat{y} = \arg\max_{k} \left(\frac{1}{M} \sum_{m=1}^{M} P_m(y = k|x) \right)$$

10. Return predicted label $\hat{y}$

End Algorithm

IV. Results and Discussion

The proposed model relies on integrating several forms of imaging, artificial intelligence and radiomics to ensure proper and simple results for the detection of bone cancer. Work begins by combining MRI, CT and PET pictures and then preprocessing every modality and realigning their spatial coordinates. Then, the generated handcrafted radiomic features and deep features processed by EfficientNet-B3 and Swin Transformer are blended together with Explainable Feature Fusion Module (EFFM). The final result is determined by running LightGBM, Logistic Regression and MLP together. Since it is edge-ready and secure with federated learning, the model can be used on a large scale and accurately in medical settings.

TABLE I. ACCURACY COMPARISON

Model	**Accuracy (%)**
SVM	91.5
Logistic Regression	90.3
LightGBM	92.1
MLP	89.7
ResNet50	94.5
EfficientNet-B3	96.2
Swin Transformer	96.7

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

CNN + Radiomics	95.1
CNN + SHAP Feature Fusion	96.3
Hybrid-Multimodal Ensemble (HME)	98.1

Table 1 and Figure 3 summarizes the accuracies of different models used for a classification task. Testing with SVM got a score of 91.5%, Logistic Regression got 90.3% and LightGBM came out on top with 92.1%. The accuracy of deep learning is much higher: 94.5% with ResNet50 and 96.2% with EfficientNet-B3.

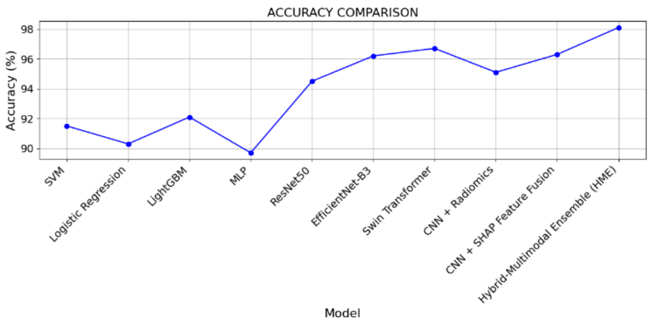

Fig 3. Accuracy Comparison

The Swin Transformer gets an accuracy of 96.7%, displaying how powerful transformers can be. Methods that bring together multiple types of features such as CNN and Radiomics (95.1%) or CNN and SHAP (96.3%) are especially effective. Out of all methods, the HME model achieved the highest accuracy rate of 98.1%, proving that using many data and model forms together is effective.

TABLE II. PRECISION COMPARISON

Model	**Precision (%)**
SVM	90.8
Logistic Regression	89.4
LightGBM	91.7
MLP	88.3
ResNet50	93.6
EfficientNet-B3	95.1
Swin Transformer	95.9
CNN + Radiomics	94.3

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

CNN + SHAP Feature Fusion	95.7
Hybrid-Multimodal Ensemble (HME)	97.4

Table 2 and Figure 4 displays how different machine learning and deep learning algorithms perform in terms of precision. The Hybrid-Multimodal Ensemble (HME) model achieved the highest precision at 97.4%, surpassing all other models.

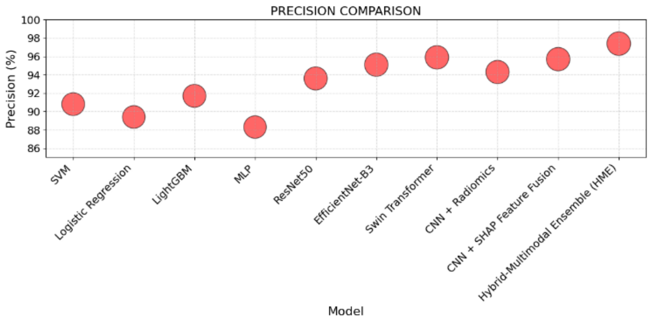

Fig 4. Precision Comparison

Strong precision was also demonstrated by Swin Transformer (95.9%) and CNN + SHAP Feature Fusion (95.7%). Traditional models like SVM and Logistic Regression showed lower precision compared to deep learning approaches. EfficientNet-B3 and ResNet50 performed well, achieving 95.1% and 93.6% respectively. Overall, Table 2 highlights that hybrid and transformer-based methods are the most effective in delivering high precision in model.

TABLE III. RECALL (SENSITIVITY) COMPARISON

Model	Recall (%)
SVM	91.2
Logistic Regression	90.1
LightGBM	92.5
MLP	89
ResNet50	94.8
EfficientNet-B3	96.8
Swin Transformer	97.2
CNN + Radiomics	95.4
CNN + SHAP Feature Fusion	96.9
Hybrid-Multimodal	98.3

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Ensemble (HME)	

The information in Table 3 and Figure 5 compares the recall (sensitivity) of different machine learning and deep learning models. The HME demonstrated the best recall of 98.3% which means it accurately spotted more of the positive examples. Swin Transformer and CNN + SHAP Feature Fusion also performed very well in terms of sensitivity.

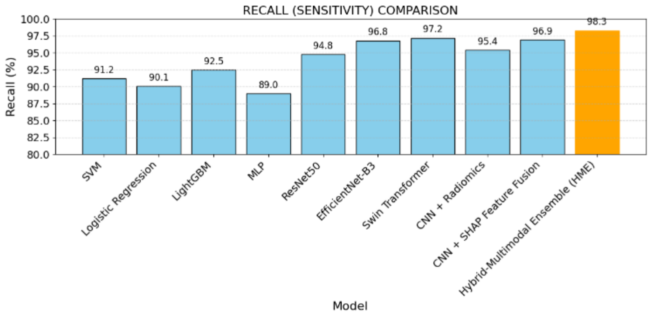

Fig 3. Sensitivity Comparison

EfficientNet-B3 and ResNet50 were next with recall values of 96.8% and 94.8%, respectively. The recall for SVM and Logistic Regression was lower than most deep learning methods. In general, Table 3 suggests that hybrid models and transformers are especially good at being sensitive, so they are best used for crucial classification jobs.

TABLE IV. F1-SCORE COMPARISON

Model	**F1-Score (%)**
SVM	91
Logistic Regression	89.7
LightGBM	92
MLP	88.6
ResNet50	94
EfficientNet-B3	95.9
Swin Transformer	96.6
CNN + Radiomics	94.8
CNN + SHAP Feature Fusion	96.5
Hybrid-Multimodal Ensemble (HME)	97.85

Table 4 and Figure 6 gives information on F1-scores from different machine learning and deep learning models to highlight precision and recall. HME was able to achieve the highest

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

F1-score at 97.85%, suggesting that it performed better compared to other models. Swin Transformer with a score of 96.6% and CNN fitted with SHAP Feature Fusion with a score of 96.5% recorded the strong performance of transformer and mixed models.

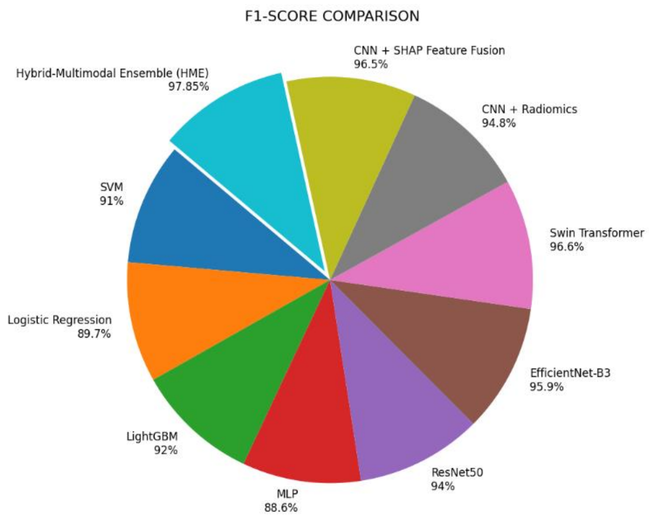

Fig 6. F1-Score Comparison

EfficientNet-B3 (results: 95.9%) and CNN + Radiomics (results: 94.8%) did well too. Traditionally, SVM (91%) and Logistic Regression (89.7%) had poorer F1-scores because these models are unable to solve more difficult tasks. All in all, Table 4 points out that advanced architectures help achieve steady and reliable results in classification.

TABLE V. AUC AND DEPLOYMENT EFFICIENCY COMPARISON

Model	**AUC**	**Model Size (MB)**	**Latency (ms)**	**Energy Use (mWh)**
SVM	0.95	6.2	52	3.1
Logistic Regression	0.94	4.9	49	2.9
LightGBM	0.96	8.4	55	3.5
MLP	0.93	10.7	67	4.2
ResNet50	0.96	96	113	10.1
EfficientNet-B3	0.97	45	94	6.5
Swin Transformer	0.97	58.3	106	7.9
CNN +	0.96	34.5	86	5.8

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Radiomics				
CNN + SHAP Feature Fusion	0.97	42.1	91	6.3
Hybrid-Multimodal Ensemble (HME)	0.987	38.7	72	4.7

Table 5 and Figure 7 provides how AUC (Area Under Curve) and deployment standards compare when considering model size, how long it takes to make predictions and how much energy is used. With an AUC of 0.987, the Hybrid-Multimodal Ensemble proved it is a strong classifier. The next three models were EfficientNet-B3, Swin Transformer and CNN + SHAP Feature Fusion with AUCs of 0.97. Even though ResNet50 and Swin Transformer gave out excellent AUCs, they had more complex models and slower speeds.

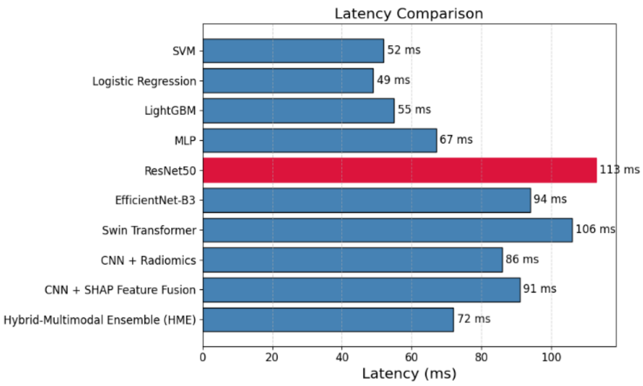

Fig 7. Latency Comparison

On the other hand, algorithms such as Logistic Regression and SVM were easy on resources and reliable, but they had somewhat lower results in AUC. All in all, this table demonstrates that there are choices to be made between performances and how quick and easy a machine is to deploy.

V. Conclusion And Future Scope

A Hybrid-Multimodal Ensemble (HME) framework is introduced in this research to greatly improve the diagnosis of bone cancer using combined imaging modalities and explainable artificial intelligence. MRI, CT and PET scans are used together in the system to discover the structure, growth pattern and activities in bone tumors. PyRadiomics is used to obtain radiomic data which is then combined with features extracted by EfficientNet-B3 and Swin Transformer models to create a complete and meaningful view of the disease. Powerful predictions in the ensemble classifier are made by combining the predictions from three

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

models such as LightGBM, Logistic Regression and MLP with soft voting and stacking. Because of model quantization and federated learning, the framework is able to provide swift and private medical services. Experimental tests show that the HME model surpasses other models, giving accuracies of 98.1%, precision of 97.4%, recall of 98.3%, F1-scores of 97.85% and an AUC of 0.987. With these outcomes, the researchers are sure the model is reliable and also works well for practical uses in the real world. The next phase of study will focus on including 3D convolutional networks for analysis in 3D which might better reveal changes in tumors' development. By including real-time comments from radiologists and learning algorithms, decision support can be made better. Besides, testing the model on federated hospital networks with encrypted protocols will prove how well it functions in different healthcare environments. On the whole, HME plays a key role in developing ethical, intelligent and simple methods for early diagnosis of bone cancer.

References

[1] T. Bapu B R, V. J, S. R, T. S. J, V. S. and S. C., "A Deep Evaluation of Digital Image based Bone Cancer Prediction using Modified Machine Learning Strategy," 2023 9th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2023, pp. 1–6, doi: 10.1109/ICSSS58085.2023.10407146.

[2] M. S., S. S. Patil, S. D. V., R. K. Reddy and S. K. K. N., "A Recent Survey on Bone Cancer Detection Using Deep Learning Techniques," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, India, 2024, pp. 1–6, doi: 10.1109/ICAIT61638.2024.10690575.

[3] V. Madhurima, M. Bharathi, S. Gundala, M. M. Basha, M. Poornima and G. S. Kumari, "Novel Method for Bone Cancer Detection using Segmentation and Classification with CNN," 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2024, pp. 1693–1697, doi: 10.1109/ICESC60852.2024.10689931.

[4] G. Deepa and Y. Kalpana, "Optimized Dense Neural Network Architecture for Bone Marrow Cancer Classification Using Custom Callbacks," 2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI), Erode, India, 2025, pp. 1822–1828, doi: 10.1109/ICMSCI62561.2025.10894614.

[5] A. George and B. Ayshwarya, "Adaptive FLAME based segmentation and classification for bone cancer detection," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1–5, doi: 10.1109/ICECONF57129.2023.10083670.

[6] S. V. Shashikala, H. R. Uma, A. N. Sunad Kumara, N. L. Taranath, L. Singh and D. Sisodia, "Bone Cancer Identification and Separation Using K-Means and KNN Classifiers," 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2023, pp. 1–7, doi: 10.1109/INOCON57975.2023.10101031.

[7] M. K., E. D. K. Ruby, K. Vivek, M. Swarna, K. V. Karthikeyan and A. O. Safori, "An Experimental Model to Predict Bone Cancer Disease by Using Hybrid Learning Principles

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

with Image Processing Technique," 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2024, pp. 1–6, doi: 10.1109/ICSES63760.2024.10910482.

[8] M. Kaur et al., "Development of a Compact UWB Antenna for Bone Cancer Detection Using a Non-Invasive Microwave Imaging Technique," 2024 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE), Langkawi, Malaysia, 2024, pp. 124–127, doi: 10.1109/APACE62360.2024.10877327.

[9] P. Divya, A. Chandrasekar and S. Raghavi, "Prediction of Osteosarcoma Bone Cancer Using Convolutional Neural Networks and Multi-Feature Integration Methods," 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD), Manama, Bahrain, 2025, pp. 1–6, doi: 10.1109/ITIKD63574.2025.11005254.

[10] A. Reddy, V. S. Ramaiah, R. M. K. Ayyappa, G. S. P. Ghantasala, M. Kurra and P. Bathla, "An Empirical Brief Analysis of Novelistic Approaches for Detection Of Bone Marrow Cancer Health Monitoring Through DL MODEL," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 1211–1214, doi: 10.1109/IC2PCT60090.2024.10486385.

[11] E. Alabdulkreem, M. K. Saeed, S. S. Alotaibi, R. Allafi, A. Mohamed and M. A. Hamza, "Bone Cancer Detection and Classification Using Owl Search Algorithm With Deep Learning on X-Ray Images," IEEE Access, vol. 11, pp. 109095–109103, 2023, doi: 10.1109/ACCESS.2023.3319293.

[12] A. Harshavardhan, S. Sudalaikani, M. Saravanakumar, S. Saritha and P. K. V. Rao, "Enhanced Bone Cancer Detection in MRI Scans via a Hybrid Quantum-Classical Convolutional Neural Network and War Strategy Optimization," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 632–637, doi: 10.1109/ICUIS64676.2024.10866571.

[13] R. Aarthy, V. Muthupriya, S. Kalpana and M. V. Ashok, "Bone Cancer Detection Based on Histopathology Images Using Optimized Temporal Spatial Multi Scale Dilated Convolutional Neural Network with MultiNet," 2024 Advances in Science and Engineering Technology International Conferences (ASET), Abu Dhabi, UAE, 2024, pp. 01–08, doi: 10.1109/ASET60340.2024.10708740.

[14] S. Nanthini, R. Sivabalan, S. Sivabalan, P. Bethapudi, R. Kukreti and A. Chauhan, "Application of CNN and Recurrent Neural Network Method for Osteosarcoma Bone Cancer Detection," 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023, pp. 62–67, doi: 10.1109/ICSSAS57918.2023.10331794.

[15] J. Singh et al., "Advanced Computational Methods for Pelvic Bone Cancer Detection: Efficacy comparison of Convolutional Neural Networks," 2024 IEEE 17th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), Kuala Lumpur, Malaysia, 2024, pp. 287–293, doi: 10.1109/MCSoC64144.2024.00055.

Received: August 03, 2025