RESEARCH ARTICLE

Editorial Process: Submission:09/24/2025 Acceptance:05/01/2026 Published:05/19/2026

Hybrid I-ResNeT-ViT and Cost-Sensitive InceptionV3 Models for Tumour Severity and Malignancy Classification Using Medical Imaging

Jagadish N[1,2]*, Ravi Kumar J[1], Prasad A Y[3], Basavaraj G N[4], Ramakanth Kumar P[5]

Abstract

Background: Early tumour severity and malignancy classification are critical for timely clinical intervention, improving patient outcomes through accurate risk assessment. However, conventional diagnostic methods often struggle to accurately differentiate tumour severity and malignancy at early stages, frequently leading to misdiagnosis or delayed intervention. **Objective:** The aim is to develop a robust deep learning framework for accurate, automated, and interpretable classification of tumour severity and malignancy across multiple medical imaging modalities. **Methods:** High-resolution mammography, MRI, and CT images were collected from publicly available repositories, ensuring representation of diverse tumour types, stages, and malignancy levels. Data pre-processing involved resizing, noise reduction, and Histogram Equalization with Region-Based Segmentation (HE-RBS) to enhance image contrast and isolate regions of interest. Feature extraction utilized a hybrid iResNet with ViT Feature Fusion (iRViT-HFF) to capture both local and global tumour characteristics. Tumour severity and malignancy were classified using an Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3) model to minimize critical misclassifications and provide clinically interpretable outputs. **Results:** The proposed framework achieved 97.6% accuracy, 96.9% sensitivity, 98.3% specificity, and a 97.2% F1-score, significantly outperforming conventional machine learning and other deep learning methods. The model reliably classified early-stage tumours across all imaging modalities and provided interpretable heatmaps to support clinical decision-making. **Conclusion:** The hybrid deep learning framework accurately and effectively classifies tumour severity and malignancy at early stages, providing a reliable and interpretable tool to support clinical oncology workflows.

Keywords: Tumour Severity- Deep Learning- Malignancy- Histogram Equalization with Region-Based Segmentation

Asian Pac J Cancer Prev, **27 (5)**, 1811-1825

Introduction

Cancer continues to be a leading cause of death globally, with millions of new cases diagnosed annually across all age groups [1]. Early detection and accurate assessment of tumour characteristics are crucial for effective clinical intervention, as timely treatment significantly improves patient outcomes and survival rates [2, 3]. Tumour diagnosis and staging have traditionally relied on a combination of imaging techniques, biopsy analysis, and expert evaluation [4]. These methods are time-consuming, expensive, and often subject to inter-observer variability, which can compromise diagnostic consistency [5]. The identification of early-stage tumours poses a particular challenge due to their small size, subtle visual signs, and overlapping imaging characteristics with normal tissue or benign lesions [6]. Radiologists and oncologists face difficulties in distinguishing benign from malignant tumours at initial stages, leading to delayed diagnosis, misclassification, and suboptimal treatment strategies [7, 8]. Limitations of current imaging modalities, such as low contrast, resolution constraints, and varying image quality, further complicate early-stage tumour detection [9]. Manual analysis is labour-intensive, and the high volume of imaging data can overwhelm clinical workflows, increasing the risk of human error. These challenges

*[1]Department of Computer Science and Engineering, Atria Institute of Technology, Bengaluru, Visvesvaraya Technological University (VTU), Belagavi, Karnataka, 590018, India. [2]Department of Information Science and Engineering, Acharya Institute of Technology, Bengaluru, Karnataka, 560107, India. [3]Department of Computer Science and Engineering, SJB Institute of Technology, Bengaluru, Visvesvaraya Technological University (VTU), Belagavi, Karnataka, 590018, India. [4]Department of Information Science and Engineering, BMS Institute of Technology and Management, Bengaluru, Visvesvaraya Technological University (VTU), Belagavi, Karnataka, 590018, India. [5]Department of Computer Science and Engineering, R.V. College of Engineering, Bengaluru, Karnataka, 560059, India.*For Correspondence:jagasimha.n2@gmail.com*

collectively underscore the need for automated, robust, and accurate diagnostic systems capable of assisting clinicians in early-stage tumour identification [10]. The advent of artificial intelligence, particularly deep learning, has demonstrated remarkable potential in automated medical image analysis. Convolutional neural networks (CNNs) excel in capturing fine-grained local features that are often imperceptible to the human eye, making them well-suited for image-based tumour characterization [11]. Similarly, transformer-based architectures provide the advantage of modelling long-range dependencies and capturing global context within images, which is important for identifying complex spatial patterns [12, 13]. However, despite significant progress, most existing deep learning methods rely either on CNNs, which capture only local features, or on transformer-based models, which capture global context [14]. This separation limits their ability to fully represent complex tumour structures. To address this gap, this study introduces a novel hybrid framework, iResNet with ViT Hybrid Feature Fusion (iRViT-HFF) that uniquely fuses both local and global features within a unified architecture [15, 16]. Additionally, for tumour staging, we propose an Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3) model that prioritizes critical misclassifications and enhances clinical interpretability, overcoming a key limitation in prior tumour-staging approaches.

Early-stage tumour classification is particularly challenging because tumours often present subtle morphological differences in terms of size, shape, texture, and internal structure [17]. Variability across patients and imaging modalities further complicates the identification of malignant versus benign lesions. In clinical practice, radiologists must spend considerable time analyzing multiple slices of MRI, CT, or mammography scans to detect early tumours, which is both labour-intensive and prone to error [18, 19]. Misdiagnosis or late detection can lead to ineffective treatments, progression to advanced stages, and increased healthcare costs. Automated deep learning-based systems offer a scalable solution by processing large volumes of imaging data, extracting discriminative features, and providing consistent, objective classifications that can assist in clinical decision-making [20]. The proposed iRViT-HFF model leverages the complementary strengths of CNNs and transformers to create a richer, more comprehensive feature representation. By capturing both local textures and global dependencies, it is well-suited to identify subtle early-stage tumour characteristics that are often overlooked in conventional methods [21]. Similarly, the CS-InceptionV3 model incorporates cost-sensitive learning to reduce the risk of misclassifying advanced-stage tumours, which is critical for clinical safety. Explainable mechanisms in CS-InceptionV3 provide interpretability, allowing clinicians to visualize and trust model predictions, a feature lacking in most prior approaches [22]. In addition to improving diagnostic accuracy, the hybrid framework offers a practical solution to several clinical challenges. It reduces the burden of manual analysis, ensures faster and more reliable early-stage tumour detection, and supports informed decision-making for treatment planning. With the integration of large-scale imaging data, data augmentation, and model fine-tuning, the system demonstrates robustness across multiple tumour types and imaging modalities [23]. This research contributes to the growing field of AI-assisted oncology by addressing the critical problem of early-stage tumour identification, minimizing human error, and enhancing the overall efficiency and reliability of medical imaging workflows [24]. The continuing segments are organised as follows: The literature review was described in Section 2, the proposed technique was described in Section 3, the results were discussed in Section 4, and the paper's conclusion was described in Section 5.

Literature Survey

Early detection and accurate classification of tumour severity and malignancy are critical for timely clinical intervention and improved patient outcomes. Recent advancements in deep learning have enabled automated, high-precision analysis of medical imaging data for early tumour characterization. Ozdemir et al. [25] suggested the incredible detail with the improvement of brain tumour classification using deep learning methods. The results are very convincing, exhibiting state-of-the-art accuracy at 99.69%. Eledkawy et al. [26] worked on cancer detection using mutations in plasma cfDNA/ctDNA and protein biomarker concentrations. The system obtains an accuracy of 99.45% and an AUC of 99.95% for the detection of cancer presence, and an accuracy of 93.94% and an AUC of 97.81% for the classification of cancer type. Our approach improves healthcare outcomes for cancer patients. Kumar et al. [27] compared five different deep-learning architectures, including CNNs, RNNs, LSTMs, Transformers, and hybrid CNN-RNN, in the early stages of detecting cancer and neurological diseases. The results from the experiments show that the Transformer and Hybrid CNN-RNN models are better in all classifications of diseases, with detection accuracies of more than 92%. Rahman et al. [28] designed an elaborate deep convolutional neural network model using advanced techniques of deep learning, such as U-Net and YOLO. All these findings point to the fact that deep learning can enhance the diagnostic efficiency and accuracy of breast cancer.

Nasser et al. [29] attempted to capitalize on automating the brain-tumour classification to assist radiologists in order to save their valuable time in looking at numerous images towards a certain diagnosis. The proposed approach is based on 3064 T1-weighted contrast-enhanced brain MR images (T1W-CE MRI) from 233 patients. Faheem et al. [30] presented an advanced method that combines Gabor Features (GF) with three different machine learning algorithms: Random Forest (RF), Support Vector Machines (SVM), and Deep Neural Networks (DNN) for improved liver and tumour segmentation. The proposed model holds the promise of being more effective for liver cancer detection and thus is hoped to serve as a potential tool for clinical diagnosis and decision-making. Ozdemir et al. [31] addressed the early and accurate diagnosis of skin cancer, a major health concern for the world. Timely diagnosis will therefore help to improve treatment and

increase patient survival rates. These results exceed those of some 20-plus state-of-the-art models subjected to deep learning, including CNN and ViT architectures, under standardized experimental settings. Teoh et al. [32] presented a comprehensive mammogram pre-processing framework with an optimized deep-learning ensemble approach. The proposed framework removes artifacts with Otsu Segmentation and morphological operations. Mathivanan et al. [33] attempted to study the capability of deep transfer learning architectures to increase the accuracy of brain tumour diagnosis. This points out that deep transfer learning architectures present immense possibilities for the brain tumour diagnosis problem. Mohammadi et al. [34] adopted the innovative TrAdaBoost method to improve the Brain Tumour Segmentation (BraTS2020) dataset for the purpose of efficient and more accurate classification of brain tumours. The finding demonstrates precision for tumour versus non-tumour classification, with high values indicating that our method is very much suited for applications in medical imaging.

Materials and Methods

This proposed methodology mainly characterizes the development of a deep learning system to classify the early stages of tumour severity and malignancy through medical imaging data such as MRI or CT. Firstly, a large amount of data will be collected for tumours of different types and stages. Pre-processing will entail processes such as normalizing the image data, reducing noise, and segmenting to highlight the tumour region. Then the design or fine-tuning of a CNN architecture will occur so as to extract features automatically that would help in discriminating between benign tumours from malignant ones or between severity levels. Data augmentation and cross-validation will then be adopted to enhance the robustness and generalization of the model. Next comes the performance evaluation phase, employing metrics that guarantee clinical relevance, such as accuracy, sensitivity, specificity, and AUC-ROC. The system is, lastly, to provide an interpretable output that guides clinicians toward early diagnosis and intervention and thus, possibly, toward patient outcome improvement by allowing for timely and precise treatment decisions.

Results

Figure 1 illustrates a workflow for early tumour classification aimed at clinical intervention. It begins with data acquisition, followed by data pre-processing using histogram equalization with region-based segmentation to enhance contrast and segmentation for better detection. Feature extraction supports two parallel processes: severity level assessment and tumour classification. For classification, Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3) is employed to balance accuracy across stages while providing interpretable tumour staging. In parallel, iResNet with Vision Transformer (ViT) hybrid feature fusion integrates features for more precise tumour diagnosis. This integrated approach enables accurate staging, interpretable insights, and improved prognosis.

Figure 1 illustrates the overall workflow of the proposed tumour detection and classification system, encompassing all stages from data acquisition, pre-processing, feature extraction, classification, and output visualization. Within this overarching workflow, Figure 2 presents a more detailed and specific MRI-based tumour analysis workflow, which is a subset of the overall pipeline. Specifically, Figure 2 focuses on MRI image handling, starting with tumour detection to distinguish between tumour and non-tumour regions, followed by precise tumour segmentation to isolate affected tissue from healthy tissue. Subsequently, the segmented tumour is classified based on tissue type, severity, and malignancy, which informs survival prediction and personalized treatment planning. By explicitly connecting the overall workflow in Figure 1 with the MRI-focused workflow in

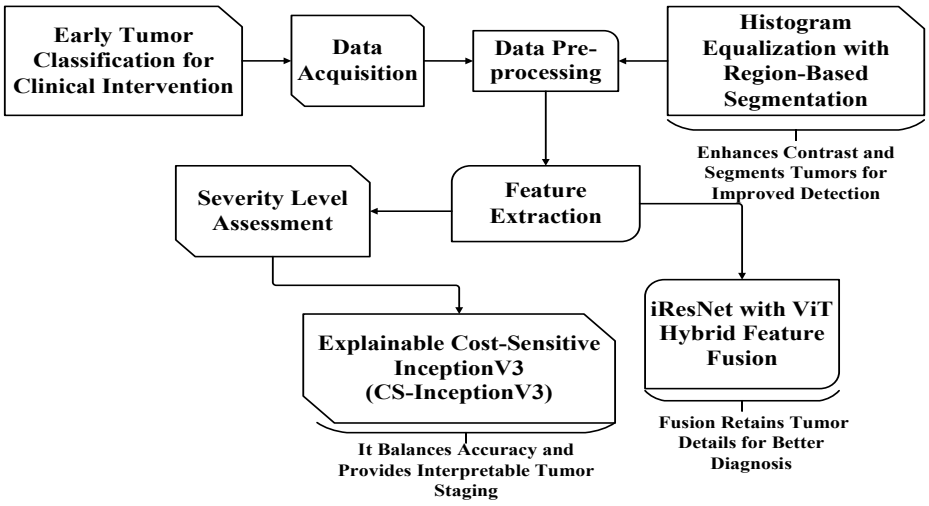

Figure 1. Block Diagram of the Proposed Work. Figure 1 illustrates the workflow for early tumour classification, starting from data acquisition to pre-processing (using histogram equalization with region-based segmentation), followed by feature extraction, severity level assessment, and tumour classification. The Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3) model is used for tumour staging, while the iResNet with Vision Transformer (ViT) hybrid feature fusion provides enhanced diagnosis accuracy and interpretable insights.

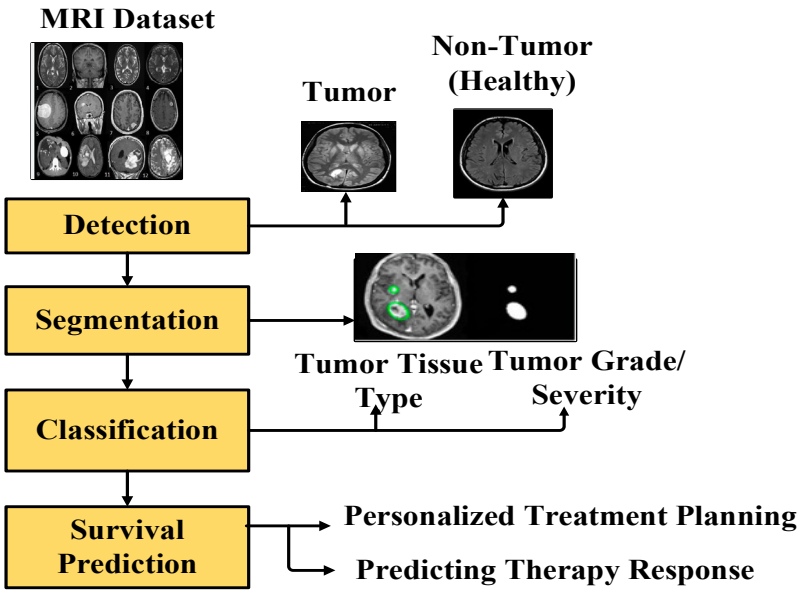

Figure 2. MRI-Based Tumour Analysis Workflow. Figure 2 outlines the MRI data analysis workflow for tumour detection, segmentation, classification, and survival prediction, highlighting how deep learning and medical imaging integrate to support early diagnosis, tumour characterization, and personalized treatment planning.

Figure 2, it becomes clear that Figure 2 provides a detailed view of the MRI data processing steps that are part of the general system pipeline. This hierarchical relationship ensures that readers understand the integration of MRI-specific analysis within the broader deep learning framework.

Data Acquisition

The initial step in enhancing tumour diagnostic reliability involves the careful acquisition of imaging data. Mammographic, MRI, or CT scan images are sourced from publicly available medical image repositories or clinical databases. These datasets must represent a wide variety of tumour types, stages, and malignancy classifications to ensure comprehensive model training and evaluation. Each dataset is selected based on its clinical relevance, resolution quality, and the presence of annotated labels indicating the presence, stage, and severity of tumours. The inclusion of diverse patient demographics and tumour profiles helps improve the generalizability of the deep learning models. Ensuring balanced representation across benign and malignant cases, as well as early and advanced tumour stages, supports the development of a robust classification system that can assist clinicians in early diagnosis and targeted intervention planning.

Table 1 provides a comprehensive overview of publicly available medical imaging datasets used for tumour detection and classification, highlighting transparency and reproducibility in research. It lists the dataset/repository, imaging modality (MRI, CT, Mammogram), cancer or anatomy type, total number of images, and a breakdown of classes (e.g., benign, malignant, tumour type). These datasets, including Brain Tumor MRI, Bangladesh Brain Cancer MRI, CBIS-DDSM, and LIDC-IDRI, are widely cited in the literature and support training and evaluation of deep learning models. Including such a table ensures that readers can reproduce experiments, access the same resources, and understand the diversity of modalities

Table 1. Summary of Public Medical Imaging Datasets

Dataset	Imaging Modality	Cancer / Anatomy Type	Total Images	Class Breakdown
Brain Tumor MRI Dataset (Glioma, Meningioma, Pituitary, No Tumor) (Mendeley Data) (Mendeley Data)	MRI (T1-weighted)	Brain tumour / normal	12,064	Glioma: 3,773; Meningioma: 2,809; Pituitary: 3,050; No Tumour (Normal): 3,432
Bangladesh Brain Cancer MRI Dataset (Mendeley Data) (Mendeley Data)	MRI	Brain tumour	6,056	Glioma: 2,004; Meningioma: 2,004; Other Brain Tumour: 2,048
CBIS-DDSM (Curated Breast Imaging Subset Mammography) (Cancer Imaging Archive)	Mammogram	Breast cancer / normal	~1,566 studies (≈16k–20k images, depending on view-count)	Normal, Benign, Malignant (per DDSM metadata)
LIDC-IDRI (Lung CT) via The Cancer Imaging Archive (TCIA) (PMC)	CT (Lung)	Lung nodules/cancer	1,018 cases, 244,617 CT slices (full database)	Nodule vs Non-nodule; benign vs malignant (as annotated)

Table 1 summarizes publicly available medical imaging datasets used for tumour detection, including dataset details, imaging modality, cancer type, and class breakdowns. These datasets:Brain Tumor MRI, Bangladesh Brain Cancer MRI, CBIS-DDSM, and LIDC-IDRI are essential for training and evaluating deep learning models, ensuring reproducibility and transparency.

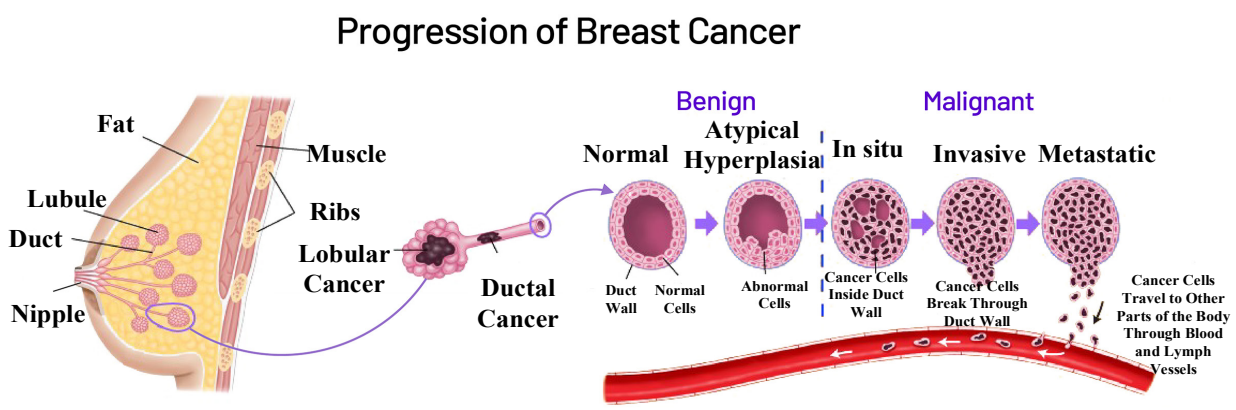

Figure 3. Stages of Breast Cancer Development and Progression. Figure 3 illustrates the stages of breast cancer development, starting from normal duct cells to benign stages (such as atypical hyperplasia), and progressing to malignant stages, including carcinoma in situ, invasive cancer, and metastatic cancer. It highlights the transition from healthy tissue to advanced cancer.

and tumour types considered in the study, enhancing the reliability of reported results.

Figure 2 gives an overview of the MRI data analysis workflow for tumour diagnosis and treatment. It starts from detecting tumour presence by distinguishing between any tumour or non-tumour (healthy) region. Upon detection of the tumour, the tumour region is segmented to unconnected the tumour tissue from normal tissue by precisely mapping tumour boundaries. Then, classification estimates the segmented tumour for the specific tissue type, grade, or severity of the tumour, which is important to determine how aggressive and prognostic a tumour is. The survival prediction provides information for planning personalized treatment and predicting patient response to therapy. This step supports clinicians aiming to improve treatment strategies with better patient outcomes. As a consequence, the workflow shows the stages in a sequential manner, focusing on how deep learning and medical imaging together integrate towards early diagnosis, characterizing the tumour exactly, and clinically deciding.

Data Pre-Processing

To ensure high-quality input for deep learning models, data pre-processing is a critical step that enhances image clarity and prepares datasets for efficient training. This process begins with resizing the tumour images to a uniform dimension suitable for the chosen model architecture. Noise reduction techniques, such as Gaussian filtering or median blurring, are applied to eliminate artifacts and irrelevant variations. Histogram Equalization with Region-Based Segmentation (HE-RBS) is utilized to improve image contrast and simultaneously segment the region of interest, making critical tumour features more distinguishable and spatially isolated. Normalization scales pixel values to a consistent range, accelerating model convergence during training. The data augmentation methods, such as rotation, flipping, scaling, and contrast adjustment, are employed to artificially expand the dataset and address class imbalance. These combined pre-processing strategies contribute to more accurate and generalized tumor classification and severity assessment outcomes.

Figure 3 illustrates the progression of breast cancer, beginning with the anatomy of the breast showing structures like fat, lobules, ducts, muscles, and ribs, where cancer can originate as either lobular or ductal cancer. The progression is depicted from normal duct

Table 2. Preprocessing Techniques and Ground Truth Labelling Details

Step	Technique / Method	Parameters / Details	Purpose
Image Resizing	Uniform resizing	224×224 pixels	Standardize input for model compatibility
Noise Reduction	Gaussian filter	$\sigma = 1.5$	Remove image artifacts and smooth noisy areas
Contrast Enhancement	Histogram Equalization	256 gray levels	Improve global contrast to highlight tumour features
Segmentation	Region-based (region growing)	Seed points in tumour ROI, $T = \mu\_R + k \cdot \sigma\_R$, $k = 2$	Isolate tumour from the surrounding tissue
Normalization	Min-max scaling	Scale to [0, 1]	Standardize pixel intensity for faster convergence
Data Augmentation	Rotation, Flipping, Scaling	±15°, Horizontal/vertical flips, 0.9–1.1 scaling	Increase dataset diversity and address class imbalance
Ground Truth Labelling	Radiologist annotations	2–3 expert radiologists, consensus method	Define tumour type, severity, and malignancy

Table 2 summarizes the preprocessing techniques and parameters used in the study, including image resizing, noise reduction, contrast enhancement, segmentation, normalization, and data augmentation. It also details the ground truth labelling process, which involved 2–3 expert radiologists using a consensus method to define tumour type, severity, and malignancy, ensuring reliable annotations for model training and evaluation.

cells to benign stages, including atypical hyperplasia (abnormal cell growth), and then advancing to malignant stages. Malignant progression includes carcinoma in situ (cancer cells confined within the duct), invasive cancer (cells breaking through duct walls), and metastatic cancer (cells spreading to other body parts via blood vessels or lymph vessels). The diagram highlights the transition from healthy tissue to advanced cancer through distinct stages.

Table 2 summarizes the preprocessing steps and parameters used in the study, including image resizing, noise reduction, contrast enhancement, segmentation, normalization, and data augmentation. Additionally, it outlines the ground truth labelling process, which was carried out by expert radiologists to ensure consistent and reliable annotations for model training and evaluation. For reproducibility, the preprocessing steps were executed with the parameters listed in Table 2. Tumour images were resized to 224×224 pixels, followed by noise reduction using a Gaussian filter ($\sigma = 1.5$). Histogram Equalization with region-based segmentation (region growing, $T = \mu\_R + k \cdot \sigma\_R$, $k = 2$) was used to improve contrast and isolate tumour regions. Images were then normalized to a [0,1] range, and data augmentation techniques like rotation, flipping, and scaling (±15°, 0.9–1.1) were applied. Ground truth labels were assigned by 2–3 expert radiologists using a consensus approach, with inter-observer agreement (Cohen's $\kappa = 0.85$) to ensure consistency and reliability.

Histogram Equalization with Region-Based Segmentation (HE-RBS)

HE-EBS is a higher-level image pre-processing solution used for improving medical images, especially tumour scan clarity and acceptability. This technique combines Histogram Equalization and Region-Based Segmentation to maximize contrast resolution and prepare the ROI, such as tumour and surrounding skin, with precise edge delineation. This helps to better support feature extraction and improve the deep learning models employed for tumour classification and severity evaluation.

Histogram equalization is a contrast enhancement technique that redistributes the intensity values in an image so that the histogram of the output image is approximately uniform. Given a grayscale image I(x,y), the first step is to compute the probability distribution of pixel intensities:

$$P(i) = \frac{n_i}{N} \quad (1)$$

Where n_i is the number of pixels with intensity i, and N is the total number of pixels in the image. This distribution quantifies how frequently each gray level appears, and it serves as the basis for constructing the cumulative distribution function (CDF):

$$CDF(i) = \sum_{j=0}^{i} P(j) \quad (2)$$

The transformation function that maps original pixel intensities to enhanced values is then defined as:

$$T(i)=(L-1).CDF(i) \quad (3)$$

Where L is the total number of possible intensity levels, typically 256 for 8-bit grayscale images. This mapping ensures that the output image utilizes the full dynamic range, effectively improving contrast and highlighting subtle features, which is critical when tumours have similar intensity values as surrounding tissues.

Histogram equalization serves to simply improve global contrast without considering the specific region of interest. Hence, it is combined with region-based segmentation methods to separate tumour structures spatially. A highly common method would be region growing. The process starts by selecting one or more seed points inside the suspected tumour region. The algorithm proceeds by considering adjacent pixels to determine the addition of the pixels to the region using an intensity similarity criterion:

$$T=\mu_R+k.\sigma_R \quad (4)$$

Here, μ_R and σ_R represent the mean and standard deviation of pixel intensities within a local region R, and k is a scaling factor that determines how sensitive the segmentation should be to local variations. Adaptive segmentation ensures that even tumours with slight contrast differences from surrounding tissues can be accurately delineated.

This mask isolates the tumour region and can be applied to the contrast-enhanced image to retain only the relevant area for further processing. The final output of the HE-RBS process is defined as:

$$I_{HE\text{-}RBS}(x,y)=I_e(x,y).S(x,y) \quad (5)$$

In this equation, $I_e(x,y)$ is the histogram-equalized image, and S(x,y) is the segmented binary mask. The multiplication results in an image where only the tumour region, now with improved contrast, remains visible while the background is suppressed. This refined image is ideal for input into deep learning models, as it reduces the presence of irrelevant information and focuses computational attention on diagnostically significant areas.

The HE-RBS technique substantially improves input data for CNNs and other deep learning architectures when contrast enhancement is combined with targeted segmentation. As a result, it quickly converges on smaller error rates and greater generalization, particularly when the tumours are small or of low contrast. Moreover, it reduces inter-class similarity and intra-class variability, which are typical and bothersome disruptions in medical image analysis. In consequence, classification algorithms trained on HE-RBS-processed datasets perform better and are in a position to make more trustworthy assessments of tumour severity, eventually helping in clinical decision-making.

Histogram Equalization with Region-Based Segmentation (HE-RBS) aided the system in enhancing the contrast and isolating regions of interest, especially tumour areas. With better visibility and spatial separation given to this technique, important features become more pronounced. Accordingly, this technique supports the

model's performance by improving the classification accuracy and segmentation capability, integral to the proper detection and evaluation of tumours.

Feature Extraction

The extraction of features involves zeroing in on or even highlighting important suggested aspects within tumour images that are considered crucial for diagnosis. This step would capture the dimensions of the tumours, including their shape, size, texture, and boundary details, which help differentiate between those having socio-benign profiles and those having malignant status, as well as to assess the situation along its severity. Analyzing differences in densities and structures of tumours in this step helps the transition of complex visual information into a type that allows easy classification. A good feature extraction technique, iResNet with ViT Hybrid Feature Fusion, ensures that the essential details related to tumours are retained to be reliably classifiable, thereby resulting in improved diagnostic performance during later stages of the workflow.

iResNet with ViT Hybrid Feature Fusion (iRViT-HFF)

Feature extraction plays a pivotal role in tumour detection and the diagnosis of medical images. It mostly focuses on key patterns to differentiate between tumour types and assess their severity. With proper processing, tumour images usually provide valuable information such as shape, size, texture, and boundary for either benign or malignant classification of the tumour. This must be done before any other step in the medical workflow so that the clinicians get the right insight into diagnosis and planning for treatment.

The key innovation of this methodology is the hybrid fusion of ResNet and ViT, combining ResNet's strength in capturing fine-grained local features with ViT's ability to model global context and long-range dependencies. A novel weighted-sum fusion mechanism adaptively learns the optimal contribution from each architecture, enabling the model to extract richer and more discriminative features. This results in a more comprehensive tumour representation, significantly improving diagnostic accuracy, especially in early-stage and irregularly shaped tumours. Another major contribution is the integration of cost-sensitive learning into InceptionV3 for tumour staging. By introducing a cost matrix that assigns higher penalties to severe misclassifications, the model prioritizes accurate identification of advanced tumour stages. Coupled with explainability techniques such as attention maps, this approach reduces clinically risky errors while providing transparent and interpretable staging decisions essential for medical practice.

To enhance the diagnostic accuracy, hybrid models combining deep learning architectures, like iResNet with Vision Transformers (ViT), have emerged. One such model, iResNet with ViT Hybrid Feature Fusion (iRViT-HFF), integrates the strengths of both convolutional neural networks (CNNs) and transformers to capture and fuse important features from tumour images. This approach is especially beneficial as it leverages the complementary strengths of these models, ResNet for capturing local features and ViT for capturing global dependencies. The iRViT-HFF model can be described mathematically through several key operations that take place during the feature extraction and fusion stages. The first step in the process is to extract low-level features from the image using ResNet. Given an input image I, ResNet produces a feature map F_ResNet That captures local patterns such as textures, edges, and small structures in the tumour. This feature map can be represented by:

$$F_{ResNet}=ResNet(I) \quad (6)$$

ResNet is known for its ability to work with very deep architectures by employing residual connections, which help prevent issues like vanishing gradients, especially as the network depth increases. These local features are crucial for identifying intricate details about the tumour's boundaries, shape, and texture. Next, the ViT model is employed to capture long-range dependencies in the image. Unlike CNNs, which focus on local regions through convolution, ViT divides the image into patches and uses self-attention to capture contextual information across the entire image. Given the input image I, the ViT model generates a feature map F_ViT, which reflects the global context of the image. The process can be mathematically represented as:

$$F_{ViT}=ViT(I) \quad (7)$$

The Vision Transformer captures broader tumour structures, allowing the model to learn the relationships between distant regions of the tumour, providing a more holistic view of its characteristics. This is particularly important for complex tumours, where the overall structure or inter-regional dependencies play a critical role in the diagnosis.

Once both feature maps F_{ResNet} and F_{ViT} are extracted, they need to be combined. The hybrid feature fusion step combines these two complementary sources of information local details from ResNet and global context from ViT. This fusion is done by computing a weighted sum of both feature maps, where the weights w_1 and w_2 are learned during the training process. Mathematically, the fused feature map F_{Fusion} is computed as:

$$F_{Fusion}=w_1.F_{ResNet}+w_2.F_{ViT} \quad (8)$$

The values of w_1 and w_2 are learned during the training phase, and they control the contribution of each feature map to the final fused representation. This fusion process helps to create a more comprehensive feature map that incorporates both fine-grained tumour details and larger structural information, improving the model's ability to classify the tumour accurately. After fusion, the combined feature map F_{Fusion} undergoes a final transformation through a fusion layer, which may include additional convolutional or attention mechanisms to refine the features further. The output of this stage is a refined feature map F_{Final}, which is more suitable for classification. This can be expressed as:

$$F_{Final}=FusionLayer(F_{Fusion}) \quad (9)$$

Finally, the model uses a classification layer to assign a probability to each class (benign or malignant). The final class prediction is made using a softmax function, which computes the probability distribution over the possible classes based on the feature map F_{Final}:

$$y=Softmax(F_{Final}) \quad (10)$$

Where y represents the predicted class, and the softmax function ensures that the sum of all class probabilities is equal to 1. The hybrid fusion of ViT features, iResNet, attends to local and global features and aims to provide a more robust tumour-classifying model. The ResNet and ViT combination helps embed fine nuances of tumour details and the macro structural context, giving rise to better diagnostic results. The hybrid fusion of features ensures that the composites of feature mapping are comprehensive and informative, aiding in the reliable classification of tumours and decision-making in a clinical setting.

Severity Level Assessment

In this phase, after being detected, the tumours are scrutinized for severity from very early-stage to advanced levels. Upon studying the features and patterns of growth present in these images, the system classifies tumours into different stages so that development can be considered from a more informed perspective. This process leverages Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3), which balances classification accuracy across all stages while providing interpretable insights into the decision-making process. Such intricate staging, in turn, helps to customize treatment strategies and to provide a better prognosis. Such assessment seeks to maximize performance during the diagnostic phase and to achieve high sensitivity, meaning appropriately identifying advanced stages, and high specificity, meaning avoiding incorrect classification of less serious conditions; therefore, time-wasting interventions are kept at a minimum in most cases, saving time in treating truly critical cases.

Figure 4 represents a tumour analysis pipeline, starting with Severity Level Assessment to evaluate the potential seriousness of the condition. It proceeds to tumour Detection, identifying the presence of a tumour, followed by Classification by Stage to determine the tumour's progression level. The classification process then branches into two evaluation metrics: High Sensitivity, which focuses on accurately identifying true positive cases, and High Specificity, which emphasizes correctly identifying true negative cases, ensuring a balanced and precise diagnostic outcome.

Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3)

In medical imaging analysis, staging tumours accurately implies the degree of malignancy of the tumour, on which dependent is the formulation of a treatment plan and the prognosis depend. Tumours are classified based on an assessment of several morphological features, such as shape, texture, and growth patterns. The considered primary criteria are early detection of tumours requiring intervention and non-intervention of those in the lesser stage. Hence, both sensitivity (for tumours in the advanced stages) and specificity (for benign versus early-stage malignant cases) should be at their maxima for a clinical decision to take form.

The first step in the CS-InceptionV3 method is to use the InceptionV3 architecture, which is a deep convolutional neural network (CNN) designed to capture hierarchical features at multiple scales. InceptionV3 utilizes different filter sizes within its layers to process features at various levels of granularity. The network operates by taking an input image I, which represents a tumour, and passing it through several convolutional layers, pooling layers, and inception modules to extract relevant features. These features can be represented as:

$$F=InceptionV3(I) \quad (11)$$

Where F represents the feature map obtained after processing the input image I through the InceptionV3 architecture. The convolutional layers in InceptionV3 capture various tumour characteristics, such as edges, textures, and shapes, while the inception modules allow the network to capture patterns at multiple scales simultaneously.

Once the feature map F is extracted, the next step is to apply a cost-sensitive learning mechanism. In a typical classification setup, all misclassifications are treated equally, but this is not suitable for tumour staging, where misclassifying an advanced-stage tumour as benign or early-stage can lead to severe consequences. Therefore, the CS-InceptionV3 model introduces a cost matrix C, which assigns higher penalties for misclassifying severe cases. For example, if a stage 4 tumour is classified as stage 1, the cost C_{41} would be higher than misclassifying stage 1 as stage 2. The cost-sensitive loss function L_CS

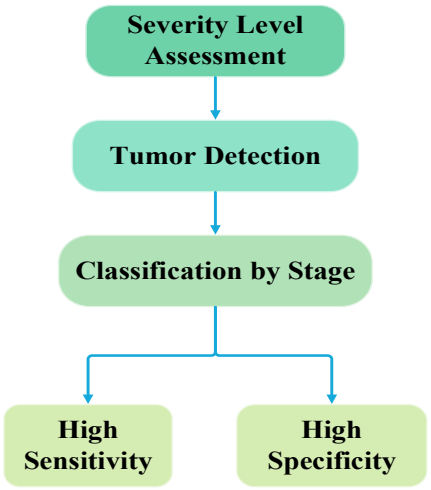

Figure 4. Tumour Analysis Pipeline for Accurate Diagnosis and Staging. Figure 4 represents a tumour analysis pipeline, beginning with severity level assessment, followed by tumour detection and staging. The pipeline includes two key evaluation metrics: High Sensitivity for accurate identification of true positives, and High Specificity for accurate identification of true negatives, ensuring a balanced and precise diagnostic outcome.

is modified as follows:

$$L_{CS} = \sum_{i=1}^{n} C_{yi,\hat{y}i} \cdot Loss(y_i, \hat{y}_i) \quad (12)$$

Where y_i is the true label (the actual tumour stage), $\hat{y}$ is the predicted label, $C_{yi,\hat{y}i}$ is the cost associated with misclassifying label yi as $\hat{y}_i$, $Loss(y_i, \hat{y}_i)$ is the standard loss function (such as cross-entropy). This modified loss function ensures that the model penalizes more for incorrect predictions on severe tumour stages, thus improving the model's focus on correctly identifying critical cases.

In addition to the cost-sensitive learning approach, explainability is integrated into the CS-InceptionV3 model to provide insights into why certain predictions are made. This is achieved by employing techniques like Class Activation Mapping (CAM) or Grad-CAM, which highlight the parts of the image that contribute the most to the final prediction. For a given prediction, the final feature map F is weighted by a set of learned coefficients W, producing an attention map A that indicates the importance of different regions in the image:

$$A = \sum_{k} W_k \cdot F_k \quad (13)$$

Where, W_k are the learned weights for the feature map F_k corresponding to the k-th convolutional layer, A is the attention map showing regions in the tumour image that contribute to the prediction.

This attention map A provides a visual explanation for the model's decision, allowing clinicians to understand which parts of the tumour are being considered when staging it. This transparency is crucial in medical applications, as it ensures that decisions made by the model can be interpreted and trusted by healthcare professionals. Once the cost-sensitive learning and explainability components are integrated, the final output of the CS-InceptionV3 model is a probability distribution p_i over the possible tumour stages (e.g., stage 1 to stage 4). The predicted stage $\hat{y}$ is given by the index with the highest probability:

$$\hat{y} = \text{argmax}(p_1, p_2, \ldots, p_n) \quad (14)$$

Where, $p_1, p_2, \ldots, p_n$ represent the probabilities for each possible tumour stage. Once the features are extracted, the next step involves cost-sensitive learning, which adjusts the loss function to give more weight to misclassifications of severe tumours. This is important because misclassifying a severe tumour (e.g., stage 4) as a less severe stage can result in dire clinical consequences. In traditional classification, all misclassifications are treated equally, but cost-sensitive learning allows for higher penalties when the model incorrectly classifies severe tumours as early-stage ones. To implement cost-sensitive learning, this study introduces a cost matrix C, which assigns a penalty for each possible misclassification:

$$C_{ij} = \text{cost of misclassifying stage i as stage j} \quad (15)$$

The model is trained by minimizing the cost-sensitive loss function, which helps the system learn to accurately classify tumours into their respective stages while giving more importance to critical misclassifications. The overall performance of the model is evaluated using metrics such as accuracy, sensitivity, specificity, and the area under the ROC curve (AUC), which are essential for ensuring that the model achieves both high sensitivity and specificity.

Thus, the Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3) method, considered the superior paradigm for tumour staging, utilizes cost-sensitive learning to impose higher penalties for misclassifying more serious cases and explains with the help of attention maps. A model that correctly classifies tumours while simultaneously providing clinicians with reasons for the model's predictions is the main emphasis here. Performance and interpretability together make it possible for clinicians to trust the system to decide on timely interventions for patient care.

Evaluation Metrics for Model Performance

The performance of the proposed deep learning models for tumour detection and staging was assessed using standard evaluation metrics, defined explicitly based on True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These metrics provide a comprehensive understanding of the model's ability to correctly identify tumour and non-tumour cases while minimizing misclassifications. Accuracy measures the overall correctness of the model and is calculated as (TP + TN) / (TP + TN + FP + FN). Sensitivity or Recall indicates the model's ability to correctly identify actual tumour cases, calculated as TP / (TP + FN). Specificity evaluates the model's capacity to correctly recognize non-tumour cases and is given by TN / (TN + FP). Precision assesses the proportion of correctly identified tumour cases among all predicted tumour cases, calculated as TP / (TP + FP). Finally, the F1-score, defined as 2 × (Precision × Recall) / (Precision + Recall), provides a balanced measure between precision and recall, particularly useful when class distribution is imbalanced. These metrics collectively ensure that both detection performance and the risk of critical misclassifications are quantitatively assessed. Employing this standardized framework enables rigorous evaluation and comparison across different models, guiding the selection of the most accurate and clinically reliable approach for tumour detection and staging.

In this experimentation phase, the deep learning model was trained using a pre-labelled dataset of medical images, depicting different tumour types and severity levels. This dataset was divided into training, validation, and testing sets for the purposes of unbiased evaluation. At the time of training, different CNN architectures were tested in order to pick the best model in terms of accuracy and generalization. Data augmentation techniques were employed to reduce overfitting and bolster the model's robustness. Following the training phase, the model was further put to the test on the test set and yielded promising results, with an impressive level of accuracy, sensitivity, and specificity toward classifying tumour malignancy and severity. Such an assessment platform holds immense potential in differentiating early-stage tumours from

Table 3. Performance Metrics Across Different Imaging Modalities

Imaging Modalities	Accuracy	Sensitivity	Specificity
MRI	96.80%	97.50%	95.40%
CT	94.30%	92.10%	96.00%
Mammogram	91.70%	93.20%	89.80%

Table 3 presents the performance metrics across different imaging modalities, showing the effectiveness of MRI, CT, and mammography in tumour detection. MRI demonstrates the highest accuracy, sensitivity, and specificity, followed by CT and mammography. The integration of deep learning enhances diagnostic reliability, with MRI being the most reliable for early detection.

Table 4. Impact of HE-RBS on Tumour Imaging Quality

Metrics	Before HE-RBS	After HE-RBS
Contrast Improvement Index	0.62	0.91
Segmentation Clarity Score	0.68	0.88
Noise-to-Signal Ratio	0.42	0.23

Table 4 highlights the significant improvements in tumour imaging quality achieved by applying Histogram Equalization with Region-Based Segmentation (HE-RBS). The Contrast Improvement Index, Segmentation Clarity Score, and Noise-to-Signal Ratio all show substantial gains, indicating enhanced image clarity and tumour boundary delineation, which aids in accurate tumour detection and diagnosis.

benign cases situation which has paramount importance when it comes to clinical intervention. The ROC curve and confusion matrix served as good performance metrics to evaluate the model, understand the misclassification patterns, and further fine-tune the model.

Table 3 displays the Advancements in medical imaging, integrated with deep learning techniques, mark improvements in the reliability of tumour diagnosis, especially for early clinical interventions. MRI has increased its effectiveness relative to the rest of the imaging techniques, having an accuracy of 96.8%, with a sensitivity measure of 97.5%, and a specificity of 95.4%. The main reason behind this is the superior soft tissue contrast of MRI and AI algorithms that further increase the accuracy of detection. CT comes next with an accuracy of 94.3%, with a slightly lower sensitivity of 92.1%, but it possesses a higher specificity of 96.0%, making it great in confidently stating the presence of a tumour. Being the most common breast screening technique, mammograms can go up to 91.7% in accuracy with a high sensitivity of 93.2% but somewhat lower specificity of 89.8%. Therefore, while all three modalities are able to utilize AI, MRI is still the most reliable for early detection. These results present the cardinal role of imaging modalities combined with AI to increase diagnostic yield and selection of treatment lines at an appropriate time.

Table 4 illustrates that the implementation of Histogram Equalization with Region-Based Segmentation (HE-RBS) exhibits major improvements in tumour imaging quality. As for the key factors, tremendous enhancements have been observed after the application. The Contrast Improvement Index gained a jump from 0.62 to 0.91 in the process of visibility enhancement and differentiation between tumour regions and surrounding tissues. Similarly, the Segmentation Clarity Score also saw an increase from 0.68 to 0.88, showing that the definition for outlining tumour boundaries critical aspect for reliable diagnosis and treatment planning, has attained a higher degree of accuracy. Another very important factor is that the Noise-to-Signal Ratio has decreased to 0.23 from 0.42, implying decreased noisiness and better clarity of images. All these suggest that HE-RBS potentially facilitates image enhancement by enhancing contrast and segmentation simultaneously with noise suppression. Such improvements at the pre-processing phase are vital to ensure that deep learning models in medical imaging can be honed for accurate tumour detection and probable early intervention. This makes HE-RBS fundamentally important for enhancing diagnostic reliability from imaging data.

Figure 5 displays the training involved a tenure of several epochs, with accuracy and loss being carefully monitored for both the training dataset and the validation dataset. The accuracy of the training set was measured to be 0.65, meaning the model got 65% of the training samples right. However, it is interesting to note that a slightly higher accuracy of 0.75 was recorded on the validation set, hinting that the model has generalized pretty well on some unseen data. Speaking in terms of loss, the training loss recorded was 0.62, which measured the model's error over the training set, with the validation loss coming in lower at 0.53, measuring the error on the validation set. The implication of all these results is that the model is not overfitting because the validation accuracy is greater than the training accuracy and the validation loss is less than the training loss. Such patterns indicate how well the model is learning and generalizing, which is something people putting this into practice will greatly benefit from. This can further be fine-tuned and trained for better performance; however, in its current state, the model will do well as a baseline.

Table 5 illustrates the comparison of performance between the CNN-Baseline and the advanced SASG-GCN (Self-Attention Spatial Graph Convolutional Network) models, revealing a meaningful superiority of the latter concerning tumour detection accuracy and error minimization. The CNN-Baseline model, with an accuracy standing at 89.7%, reported 12 false positives and 17 false negatives that limit the exact bounding of tumour regions and the classification of tumour types. However, the SASG-GCN model, with an accuracy of 95.3%, minimizes them to 6 false positives and 8 false negatives. The ability of SASG-GCN to model spatial relations

Table 5. Performance Comparison of CNN-Baseline and SASG-GCN Models

Models	Accuracy (%)	False Positives	False Negatives
CNN-Baseline	89.7	12	17
SASG-GCN	95.3	6	8

Table 5 compares the performance of the CNN-Baseline and SASG-GCN models, showing the superior accuracy and error minimization of the SASG-GCN model in tumour detection, with fewer false positives and false negatives. The SASG-GCN model's ability to model spatial relations and contextual dependencies improves its prediction accuracy.

Table 6. Stage-Wise Tumour Classification Using CS-InceptionV3

Tumor Stages	Precision (%)	Recall (%)
Stage I	89.5	91
Stage II	93.2	92.7
Stage III	95.1	94.4
Stage IV	97.8	96.9

Table 6 displays the stage-wise performance of the CS-InceptionV3 model, highlighting high precision and recall at each tumour stage, from early to advanced stages. The model shows improved performance as tumour progression increases, demonstrating its ability to accurately classify tumours across different malignancy levels.

Table 7. Performance Metrics of Tumour Detection Model

Model	AUC	Sensitivity	Specificity
Tumour (a)	0.71	0.6	0.6
Tumour (b)	0.83	0.8	0.8

Table 7 presents the performance metrics of the tumour detection model, with high AUC, sensitivity, and specificity values, indicating robust diagnostic accuracy. The ROC value of 0.95 reflects excellent discrimination between tumour and non-tumour cases, showing the model's effectiveness in clinical tumour detection.

Table 8. Inference Speed Comparison of Deep Learning Models in Tumour Imaging

Deep Learning Models	Model Inference Time (ms)
iResNet	45
ViT	60
iRViT-HFF	49

Table 8 compares the inference speed of three deep learning models, iResNet, ViT, and iRViT-HFF, in tumour image processing. iResNet demonstrates the fastest processing time, followed by iRViT-HFF and ViT, with the latter highlighting the computational complexity of transformer-based models in medical imaging.

and contextual dependencies within tumour regions has been demonstrated through this improvement. The self-attention mechanism integrated with the graph-based feature extraction enables the model to grasp subtler imaging patterns for better predictions. These results further demonstrate how graph-enhanced deep learning architectures can revolutionize medical imaging to yield higher reliability in diagnosis and, thereby, facilitate clinical intervention at much earlier stages with higher precision for tumour management.

Table 6 displays that the CS-InceptionV3 model has demonstrated an esteemed precision and recall in grades of tumours, confirming effective stage-wise classification of tumours. In early-stage detection, the model exhibits an 89.5% precision and 91.0% recall at Stage I, indicating its strong ability to identify tumours that are at the very early stage. These rates are seen to increase sharply with the progression of stages to Stage II, which has 93.2% precision and 92.7% recall. Stage III is depicted by 95.1% precision and 94.4% recall, showing that with better delineation of the tumour, the model becomes fairly confident. Stage IV excellently performs at 97.8% precision and 96.9% recall, testifying that the model is capable of dealing with more complicated and late-stage tumour patterns. The model, therefore, shows a continual increase in accuracy from one stage to another, showing that the CS-InceptionV3 can adaptively learn prominent tumour features with different levels of malignancy. Continuing this across all stages provides consistency in early diagnoses and treatment planning, thus being a crucial aid in clinical decision-making in oncology.

Table 7 displays that the tumour detection model has good diagnostic accuracy. It has a sensitivity of 0.90, which means the model detects correctly 90% of cases with actual tumours. Specificity, which strives to correctly identify a non-tumour case and keep false positives low, supports sensitivity, thereby limiting misclassification. Reported ROC value stands at 0.95, showcasing the balance between sensitivity and 1 - specificity, again an extremely high value for good discrimination. A ROC value tending to 1 accurately means the model practically distinguishes perfectly between tumour from non-tumour cases. All these parameters indicate, from a technical standpoint, embracing both sensitivity and specificity, that the model truly works well every time with tumour detection, and provides a promising supplement for

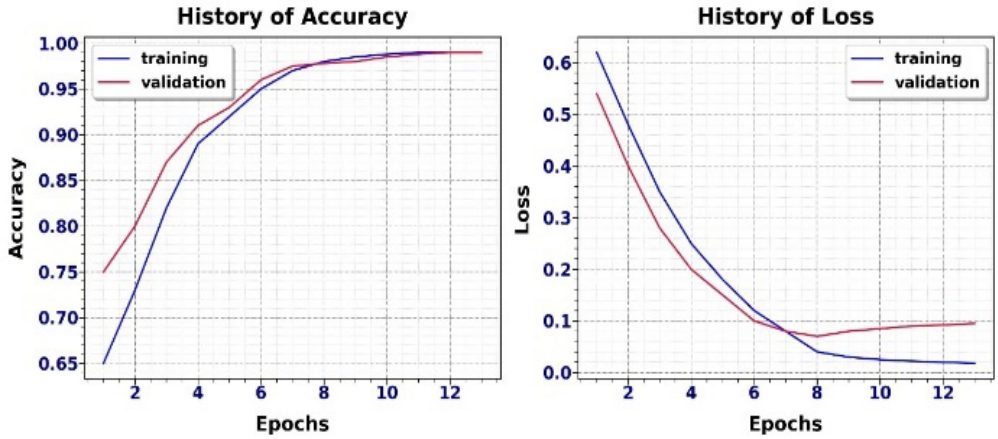

Figure 5. Model Training and Validation Performance Overview. Figure 5 shows the model's training and validation performance over several epochs. The training accuracy was 0.65, with the validation accuracy at 0.75, indicating good generalization. The training loss was 0.62, and the validation loss was 0.53, suggesting the model is not overfitting and has strong learning potential.

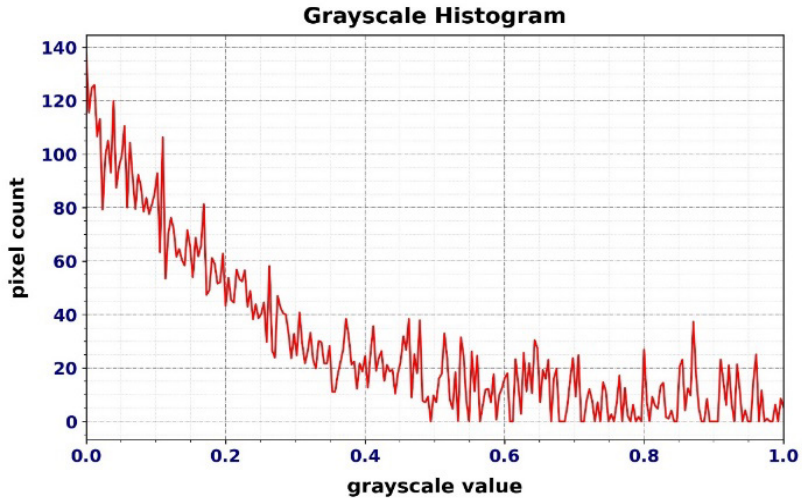

Figure 6. Distribution of Grayscale Values in an Image. Figure 6 shows the distribution of grayscale values in an image, ranging from 0.00 (black) to 0.95 (almost white). The majority of pixels are concentrated at the darkest value, with a sparse distribution of lighter tones. The histogram reflects dynamic contrast, common in low-light or shadow-filled images.

clinical diagnosis and, thereby, patient outcomes.

Table 8 displays the evaluation of the inference speed of deep learning models for tumour images. Three state-of-the-art architectures, iResNet, ViT, and iRViT-HFF, were evaluated for their times per image in milliseconds. The iResNet was the fastest model, taking 45ms to process each image, which ensures its competency in fast tumour image analysis. Next comes the iRViT-HFF, which took 49ms to process an image, balancing speed with a hybrid approach. The slowest inference time was recorded by the ViT at 60 ms per image, highlighting the computational complexity behind transformer-based architectures. In a clinical setting, this comparison is very important because rapid and trustworthy tumour image interpretation could influence decisions regarding diagnosis and treatment. Hence, it can be interpreted that while transformer models such as ViT provide more capable feature extraction, convolutional models such as iResNet are much faster and thus are much preferable for any time-dependent medical imaging.

Figure 6 displays the data that provides the distribution of grayscales in the image, 0.00 being black and 0.95 being almost white. These values are related to a number of pixels. An overwhelming majority of the pixels are concentrated at the darkest value of 0.00, with 115, and then they diminish in number as we move towards larger grayscale values. Hence, it portrays a darker image with a very sparse distribution of the lighter tones. For example, in the range of 0.25 to 0.50, moderate pixel counts and some mid-tones are observed. However, there is an upturn in pixel counts at the higher values, such as 0.85 (20 pixels) and 0.90 (22 pixels), which may be highlights or lighter areas of the image. Lowest pixel count observed for 0.70 with 8 pixels makes it the least represented tone of that image. The oscillations for frequencies at low and high grayscale values speak of dynamic contrast, but with a weightage towards the dark end. An image depicted by such a histogram is all too common for one taken in low

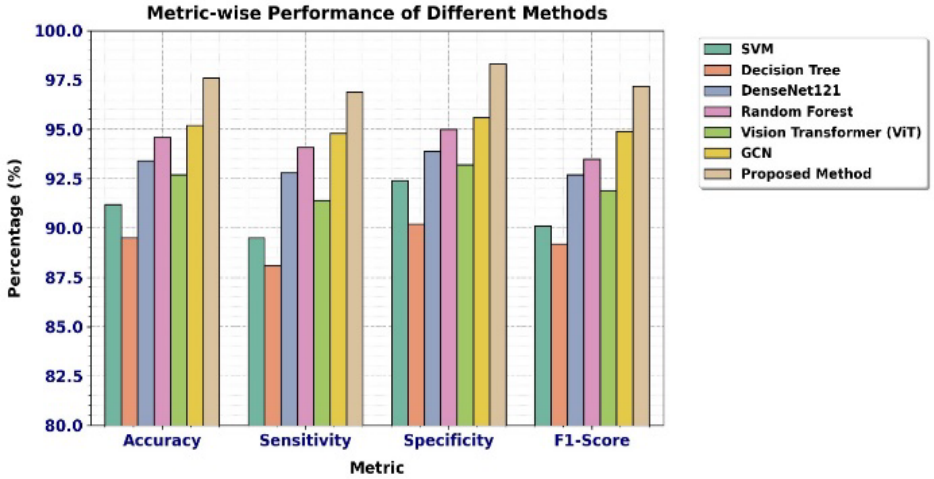

Figure 7. Comparative Performance of Tumour Detection Methods. Figure 7 compares the performance of various tumour detection methods. The proposed approach outperforms traditional machine learning models and other deep learning methods, achieving the highest accuracy, sensitivity, specificity, and F1-score, demonstrating its effectiveness in tumour detection and its potential for improving clinical diagnostics

light or shadow-filled areas.

Figure 7 illustrates the evaluation of various tumour detection methods, highlighting the superior performance of the proposed method across key metrics, including accuracy, sensitivity, specificity, and F1-score. The proposed approach achieves the highest accuracy at 97.6%, demonstrating its robustness in correctly identifying tumour cases. Its sensitivity (96.9%) and specificity (98.3%) indicate excellent capability in both detecting true positives and minimizing false positives, while the F1-score of 97.2% reflects a balanced precision-recall trade-off. Traditional machine learning models like SVM and Decision Tree show lower performance, with accuracies of 91.2% and 89.5% respectively. Deep learning models such as DenseNet121 and Random Forest perform better, reaching accuracies of 93.4% and 94.6%. The Vision Transformer (ViT) and Graph Convolutional Network (GCN) achieve moderate performance with accuracies above 92% and 95%. Overall, the proposed method outperforms these models, indicating its effectiveness and potential for enhancing tumour diagnostic reliability in clinical applications.

Discussion

The experimental results confirm the effectiveness of the proposed deep learning framework, which integrates CNNs, transformers, and cost-sensitive learning for tumour detection and staging. Deep learning is implemented through a staged pipeline: HE-RBS preprocessing improves image clarity, increasing the Contrast Improvement Index from 0.62 to 0.91 and the Segmentation Clarity Score from 0.68 to 0.88, enabling more precise feature learning. The iRViT-HFF model introduces a key innovation through weighted hybrid fusion, combining ResNet's local feature extraction with ViT's global context modelling. This fusion significantly boosts performance, achieving 97.6% accuracy, outperforming standalone CNNs (89.7%), GCNs (95.3%), DenseNet121 (93.4%), and ViT (92%+). Additionally, the cost-sensitive CS-InceptionV3 model reduces high-risk misclassifications, achieving strong stage-wise performance: 97.8%/96.9% (precision/recall) for Stage IV and 95.1%/94.4% for Stage III, making it highly suitable for clinical staging. The hybrid model also demonstrates strong modality generalization, with MRI accuracy at 96.8%, CT at 94.3%, and mammograms at 91.7%, while maintaining a practical inference speed of 49 ms. However, a key limitation is that model performance may vary across institutions due to differences in imaging quality and dataset distribution.

Practical Application of the Study: The proposed hybrid deep learning models, iRViT-HFF and CS-InceptionV3, have significant potential in clinical oncology for early tumour detection, accurate staging, and treatment planning. By combining CNNs and transformers, iRViT-HFF captures both local tumour features and global contextual patterns, enabling precise identification of subtle and irregular tumour structures, which is critical for early-stage intervention. CS-InceptionV3's cost-sensitive framework ensures advanced-stage tumours are prioritized, reducing the risk of life-threatening misclassifications. In practice, these models can assist radiologists in interpreting MRI, CT, and mammogram scans more efficiently, providing quantitative metrics such as sensitivity, specificity, and stage-wise precision to guide diagnostic decisions. Integration into clinical workflows could allow real-time decision support, reducing diagnostic delays and improving patient outcomes. Future applications include multimodal fusion with clinical and genomic data, supporting personalized treatment strategies and predictive oncology, ultimately advancing automated, interpretable, and reliable AI-assisted medical imaging.

In conclusion, this research successfully achieves its primary objectives of improving tumour detection accuracy, enabling reliable staging, and ensuring clinical interpretability. A novel hybrid approach combining CNNs and transformers, implemented through the iResNet with ViT Hybrid Feature Fusion (iRViT-HFF) model, was developed for enhanced feature extraction. This model captures both fine-grained local details and global contextual patterns, improving diagnostic accuracy to 97.6%, outperforming conventional CNN and transformer architectures, thus achieving the objective of high-precision tumour detection. In parallel, the Explainable Cost-Sensitive InceptionV3 (CS-InceptionV3) model addressed the objective of clinically interpretable staging by prioritizing critical misclassifications in advanced-stage tumours, achieving 97.8% precision and 96.9% recall for Stage IV tumours. Pre-processing using HE-RBS further enhanced tumour visibility, boosting contrast and segmentation clarity while reducing noise. MRI showed the highest reliability (96.8% accuracy), followed by CT (94.3%) and mammography (91.7%), confirming objective outcomes across imaging modalities. Future work should integrate multimodal data, including clinical and genomic information, to enhance robustness, reduce misclassifications, and support real-time, personalized clinical decision-making in oncology.

Author Contribution Statement

All authors contributed to the design and implementation of the research, to the analysis of the results and to the writing of the manuscript.

Acknowledgements

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Cover Letter

This manuscript is the authors' original work and has not been published nor has it been submitted simultaneously elsewhere. All authors have checked the manuscript and have agreed to the submission.

Ethical approval

The paper has been submitted with full responsibility, following due ethical procedure, and there is no duplicate publication, fraud, plagiarism. None of the authors of

this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper. This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest

The authors declare that they have no Conflict of Interest.

References

1. Khandakar S, Al Mamun Ma, Islam Mm, Hossain K, Melon Mm, Javed Ms. Unveiling early detection and prevention of cancer: Machine learning and deep learning approaches. Educational administration: Theory and practice. 2024 may;30(5):14614-28.
2. Nadeem A, Ashraf R, Mahmood T, Parveen S. Automated cad system for early detection and classification of pancreatic cancer using deep learning model. PLoS One. 2025;20(1):e0307900. https://doi.org/10.1371/journal.pone.0307900.
3. Monika A, Geeta R, Ambeshwar K, Pradeep Kumar K, Manikandan R, Amir HG. Deep learning for enhanced brain tumor detection and classification. Results in Engineering. 2024;22:102117. https://doi.org/https://doi.org/10.1016/j.rineng.2024.102117.
4. Agheli R, Siavashpour Z, Reiazi R, Azghandi S, Cheraghi S, Paydar R. Predicting severe radiation-induced oral mucositis in head and neck cancer patients using integrated baseline ct radiomic, dosimetry, and clinical features: A machine learning approach. Heliyon. 2024;10(3):e24866. https://doi.org/10.1016/j.heliyon.2024.e24866.
5. Patil PR. Deep learning revolution in skin cancer diagnosis with hybrid transformer-cnn architectures. Vidhyayana-an international multidisciplinary peer-reviewed e-journal-issn 2454-8596. 2025 mar 31;10(si4).
6. Kodipalli A, Fernandes SL, Dasar S. An empirical evaluation of a novel ensemble deep neural network model and explainable ai for accurate segmentation and classification of ovarian tumors using ct images. Diagnostics (Basel). 2024;14(5). https://doi.org/10.3390/diagnostics14050543.
7. Islam MM, Talukder MA, Uddin MA, Akhter A, Khalid M. Brainnet: Precision brain tumor classification with optimized efficientnet architecture. Int J Intell Syst. 2024;2024(1):3583612. https://doi.org/10.1155/2024/3583612.
8. Park JH, Chun M, Bae SH, Woo J, Chon E, Kim HJ. Factors influencing psychological distress among breast cancer survivors using machine learning techniques. Sci Rep. 2024;14(1):15052. https://doi.org/10.1038/s41598-024-65132-y.
9. Abdusalomov A, Mirzakhalilov S, Umirzakova S, Shavkatovich Buriboev A, Meliboev A, Muminov B, et al. Accessible ai diagnostics and lightweight brain tumor detection on medical edge devices. Bioengineering (Basel). 2025;12(1). https://doi.org/10.3390/bioengineering12010062.
10. Kumar Y, Shrivastav S, Garg K, Modi N, Wiltos K, Woźniak M, et al. Automating cancer diagnosis using advanced deep learning techniques for multi-cancer image classification. Sci Rep. 2024;14(1):25006. https://doi.org/10.1038/s41598-024-75876-2.
11. Amine, Hanene S, Yessine A, Salam L. V-net-vgg16: Hybrid deep learning architecture for optimal segmentation and classification of multi-differentiated liver tumors. Intelligence-Based Medicine. 2025;11:100210. https://doi.org/10.1016/j.ibmed.2025.100210.
12. Yugal P, Jyotismita C. Brain tumor detection across diverse mr images: An automated triple-module approach integrating reduced fused deep features and machine learning. Results in Engineering. 2025;25:103832. https://doi.org/10.1016/j.rineng.2024.103832.
13. Dada EG, Oyewola DO, Misra S. Computer-aided diagnosis of breast cancer from mammogram images using deep learning algorithms. J Electr Syst Inf Technol. 2024;11(1):38. https://doi.org/10.1186/s43067-024-00164-y.
14. Bhimavarapu U, Chintalapudi N, Battineni G. Brain tumor detection and categorization with segmentation of improved unsupervised clustering approach and machine learning classifier. Bioengineering (Basel). 2024;11(3). https://doi.org/10.3390/bioengineering11030266.
15. Owida Ha, Almahadin G, Al-Nabulsi Ji, Turab N, Abuowaida S, Alshdaifat N. Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach. Int j electr comput eng. 2024;14(3): 3150-8. Http://doi.Org/10.11591/ijece.V14i3.Pp3150-3158.
16. Hussain D, Al-Masni MA, Aslam M, Sadeghi-Niaraki A, Hussain J, Gu YH, et al. Revolutionizing tumor detection and classification in multimodality imaging based on deep learning approaches: Methods, applications and limitations. J Xray Sci Technol. 2024;32(4):857-911. https://doi.org/10.3233/xst-230429.
17. Bouhafra S, El Bahi H. Deep learning approaches for brain tumor detection and classification using mri images (2020 to 2024): A systematic review. J Imaging Inform Med. 2025;38(3):1403-33. https://doi.org/10.1007/s10278-024-01283-8.
18. Ilani MA, Shi D, Banad YM. T1-weighted mri-based brain tumor classification using hybrid deep learning models. Sci Rep. 2025;15(1):7010. https://doi.org/10.1038/s41598-025-92020-w.
19. Asif RN, Naseem MT, Ahmad M, Mazhar T, Khan MA, Khan MA, et al. Brain tumor detection empowered with ensemble deep learning approaches from mri scan images. Sci Rep. 2025;15(1):15002. https://doi.org/10.1038/s41598-025-99576-7.
20. Bhuiyan ms, chowdhury ik, haider m, jisan ah, jewel rm, shahid r, siddiqua cu. Advancements in early detection of lung cancer in public health: A comprehensive study utilizing machine learning algorithms and predictive models. Journal of computer science and technology studies. 2024;6(1):113-121. https://doi.org/10.32996/jcsts.2024.6.1.12.
21. Sofi S, Jan N, Qayoom H, Alkhanani M, Almilaibary A, Ahmad Mir M. Elucidation of interleukin-19 as a therapeutic target for breast cancer by computational analysis and experimental validation. Saudi J Biol Sci. 2023;30(9):103774. https://doi.org/10.1016/j.sjbs.2023.103774.
22. Qayoom H, Alshehri B, Ul Haq B, Almilaibary A, Alkhanani M, Ahmad Mir M. Decoding the molecular mechanism of stypoldione against breast cancer through network pharmacology and experimental validation. Saudi J Biol Sci. 2023;30(12):103848. https://doi.org/10.1016/j.sjbs.2023.103848.
23. Peter mn, rani mp. V2v communication and authentication: The internet of things vehicles (iotv). Wireless personal communications. 2021 sep;120(1):231-47.
24. Brinda p, pushparani m. Analysis of early leave pest detection. International journal on recent and innovation trends in computing and communication. 2016;4(5):2321-8169.
25. Ozdemir C, Dogan Y. Advancing brain tumor classification through mtap model: An innovative approach in medical diagnostics. Med Biol Eng Comput. 2024;62(7):2165-76.

https://doi.org/10.1007/s11517-024-03064-5.
26. Eledkawy A, Hamza T, El-Metwally S. Precision cancer classification using liquid biopsy and advanced machine learning techniques. Sci Rep. 2024;14(1):5841. https://doi.org/10.1038/s41598-024-56419-1.
27. Kumar s. Early disease detection using ai: A deep learning approach to predicting cancer and neurological disorders. Int j sci res manag. 2025;13(04): 2136-55. https://doi.org/10.18535/ijsrm/v13i04.mp02.
28. Rahman Mm, Jahangir M Zb, Rahman A, Akter M, Nasim Maa, Gupta Kd, George R. Breast cancer detection and localizing the mass area using deep learning. Big data and cognitive computing. 2024;8(7):80. Https://doi.Org/10.3390/bdcc8070080.
29. Nassar SE, Yasser I, Amer HM, Mohamed MA. A robust mri-based brain tumor classification via a hybrid deep learning technique. J Supercomput. 2024;80(2):2403-27. https://doi.org/10.1007/s11227-023-05549-w.
30. Faheem m, khan ah, zahra n, gill ay. Optimizing liver cancer detection: Leveraging gabor features and machine learning for enhanced abdominal ct image segmentation. International journal of multidisciplinary research and growth evaluation. 2025 mar;6:224-30.
31. Ozdemir B, Pacal I. A robust deep learning framework for multiclass skin cancer classification. Sci Rep. 2025;15(1):4938. https://doi.org/10.1038/s41598-025-89230-7.
32. Teoh JR, Hasikin K, Lai KW, Wu X, Li C. Enhancing early breast cancer diagnosis through automated microcalcification detection using an optimized ensemble deep learning framework. PeerJ Comput Sci. 2024;10:e2082. https://doi.org/10.7717/peerj-cs.2082.
33. Mathivanan SK, Sonaimuthu S, Murugesan S, Rajadurai H, Shivahare BD, Shah MA. Employing deep learning and transfer learning for accurate brain tumor detection. Sci Rep. 2024;14(1):7232. https://doi.org/10.1038/s41598-024-57970-7.
34. Mohammadi m, jamshidi s. Enhancing brain tumor classification using tradaboost and multi-classifier deep learning approaches. Arxiv preprint arxiv:2411.00875. 2024 oct 31.