



Article

Automated Differentiation of Hepatic Cysts and Metastatic Tumors Using a Deep Learning Ensemble Framework on CT Imaging in Low-Resource Areas

Mamoun Qjidaa ¹, Amine Souadka ², Anass Benfares ³, Mohammed Amine El Azami El Hassani ², Amine Benkabbou ², Anass Majbar ², Zakaria El Moatassim ², Maroua Oumlaz ⁴ , Oumayma Lahnaoui ², Raouf Mouhcine ², Hassan Qjidaa ^{5,*}, Ahmed Lakhssassi ⁴, Ouazzani Jamil Mohammed ⁵ and Abdeljabbar Cherkaoui ¹

¹ Higher School of Applied Sciences of Tangier, Abdelmalek Essaadi University, Tangier 93030, Morocco; qmamoun@gmail.com (M.Q.)

² National Institute of Oncology of Rabat, Mohammed V University, Rabat 10000, Morocco; amine.souadka@um5.ac.ma (A.S.); mohammedamine.elazamielhassani@um5.ac.ma (M.A.E.A.E.H.); amine.benkabbou@um5.ac.ma (A.B.); anass.majbar@um5.ac.ma (A.M.); zakaria.elmoatassim@um5.ac.ma (Z.E.M.); oumay-ma.lahnaoui@um5.ac.ma (O.L.); raouf.mouhcine@um5.ac.ma (R.M.)

³ Faculty of Sciences Dhar le Mehraz of Fez, Sidi Mohammed Ben Abdellah University, Fez 30000, Morocco; anas.benfraess@usmba.ac.ma

⁴ Department of Computer Science and Engineering, Université du Québec en Outaouais, Gatineau, QC J8X 3X7, Canada; oumm08@uqo.ca (M.O.)

⁵ Faculty of Engineering Sciences, Private University of Fez, Fez 30999, Morocco; ouazzani@upf.ac.ma

* Correspondence: hassan.qjidaa@usmba.ac.ma

Abstract

Background/Objectives: In resource-limited settings, where access to advanced imaging modalities such as magnetic resonance imaging (MRI) and histopathological confirmation may be limited, differentiating between hepatic cysts and metastatic lesions based solely on computed tomography (CT) remains challenging. This limitation may affect diagnostic confidence and increase the risk of misclassification, potentially impacting clinical decision making and patient management. In this study, we aimed to explore a more direct and automated approach for classifying hepatic lesions from CT images. **Methods:** We developed a deep learning-based framework combining transfer learning, decision fusion, and a stacking strategy by integrating five CNN architectures. The study included 100 patients, equally divided between metastatic liver tumors and pathological hepatic cysts. The dataset was built from both public data (LiTS) and internal clinical cases, and then split into training and testing sets. **Results:** The proposed stacking model provided the most consistent results, reaching an accuracy of 0.98, with high precision and sensitivity. The improvements in individual models, although moderate, were observed across all evaluation metrics. **Conclusions:** Overall, this approach offers a practical and reliable way to classify hepatic lesions with minimal manual intervention. It may help improve consistency in diagnosis and could be integrated into clinical workflows to support decision making in low-resource areas.

Keywords: hepatic lesions; liver metastases; hepatic cysts; computed tomography; deep learning; convolutional neural networks; transfer learning; ensemble learning; stacking model; decision level fusion; computer aided diagnosis; medical imaging; low resource settings; automated classification; clinical decision support



Academic Editor: Hiroyuki Yoshida

Received: 8 May 2026

Revised: 14 June 2026

Accepted: 18 June 2026

Published: 7 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

According to recent estimates from the GLOBOCAN 2022 database, liver cancer is the sixth most frequently diagnosed cancer worldwide, with approximately 866,000 new cases and nearly 759,000 deaths, making it the third leading cause of cancer-related mortality and accounting for nearly 8% of all cancer deaths [1].

In this context, machine learning (ML) has emerged as a powerful tool for extracting and modeling complex hepatic patterns from medical imaging data. These methods integrate information related to intensity, texture, shape, and multiphasic dynamics to automatically identify discriminative signatures that are often imperceptible to visual inspection.

Despite the clear potential of differentiating hepatic cysts from tumors using radiomics-based machine learning approaches, this strategy presents significant limitations in terms of predictive accuracy. These limitations are often associated with human-induced segmentation errors during the annotation of internal datasets, variability in radiomic features, and their sensitivity to CT acquisition parameters. Such constraints motivate the exploration of more advanced deep learning-based methods capable of bypassing the manual segmentation step.

Furthermore, the clinical translation of radiomics models remains limited, mainly due to the high dimensionality of features, which may lead to overfitting, reduced robustness, and poor generalizability across different datasets and imaging protocols.

Although recent methodological advances have improved performance, several limitations persist. Despite promising results, many artificial intelligence (AI)-based approaches in hepatic oncology remain insufficiently integrated into clinical practice due to limited external validation, lack of generalizability, and insufficient robustness of performance metrics, thereby hindering their reliable deployment in real-world settings [2–9].

These limitations can be primarily attributed to three major challenges: (i) errors arising from manual annotation of training datasets, particularly in tumor segmentation; (ii) limited reliability and consistency of model predictions; and (iii) the high computational and economic cost associated with complex architectures, especially for exhaustive radiomic feature extraction from volumetric tumor regions of interest.

The main objective of this study is to develop a non-invasive, cost-effective, and highly accurate framework for differentiating hepatic cysts from metastatic liver tumors, with the aim of enhancing clinician confidence and facilitating clinical validation. The non-invasive nature of this approach relies on the implementation of an automated deep learning method based on computed tomography (CT) imaging.

In this framework, we propose the development of a deep learning (DL) model as a promising alternative. CNNs are capable of automatically learning complex and discriminative representations directly from raw images, without requiring manual feature extraction steps. This approach enables the capture of subtle patterns, sometimes imperceptible to the human eye or conventional radiomics methods, and offers improved robustness and generalizability in predicting the nature of hepatic lesions (cyst vs. tumor).

To further enhance model performance, annotation-related biases were reduced by eliminating the need for manual segmentation. In addition, a transfer learning-based deep learning framework was developed, combining transfer learning, decision fusion, and a stacking strategy by integrating five CNN architectures. The final classification is achieved using a majority voting scheme, thereby improving the robustness and reliability of the model predictions.

The main innovation of this study lies in the proposed decision fusion strategy, which integrates the complementary feature representations learned by five pretrained CNN architectures through transfer learning and a stacking-based ensemble mechanism. This approach improves the robustness and generalization of individual CNN models while

providing an efficient alternative to ViT-based methods, which typically require large-scale datasets and high computational resources.

2. Literature Review

Given the significant differences in prognosis and treatment strategies, accurate differentiation between hepatic cysts and liver tumors is essential for effective clinical management. Several studies have addressed this issue. Chenin et al. emphasized that a thorough understanding of imaging patterns is crucial for distinguishing benign cysts from other hepatic lesions requiring treatment. They reported that hepatic cysts typically exhibit specific radiological features, including well-defined lesions with homogeneous fluid content and no contrast enhancement. However, the presence of thickened walls, septations, or solid components should raise suspicion of a potentially non-benign lesion, highlighting the importance of accurate differentiation from other hepatic abnormalities to ensure appropriate management [10].

Similarly, Mavilia et al. [11] highlighted that the differentiation of cystic lesions is critical for distinguishing benign from malignant conditions, and that medical imaging plays a central role in this decision-making process.

Wang et al. proposed approaches based on automated image analysis and artificial intelligence to differentiate between benign and malignant hepatic lesions. They developed a radiomics model based on multiphase computed tomography (CT), demonstrating high performance for reliable and non-invasive classification of liver lesions [12].

Granata et al. investigated imaging strategies aimed at improving the identification of malignant hepatic lesions, highlighting the complementary roles of ultrasound, computed tomography, magnetic resonance imaging (MRI), and hybrid techniques such as PET/CT, as well as their impact on clinical management [13].

Alshomrani et al. [14] emphasized that recent advances in MRI, including multiparametric imaging, hepatospecific contrast agents, and the integration of artificial intelligence, have significantly improved the detection and characterization of hepatic lesions. These innovations enable more accurate diagnosis and contribute to improved therapeutic management. However, several limitations remain, including high cost, limited accessibility, variability in acquisition protocols, and the need for standardization of radiomics methods. The authors suggested future directions involving broader integration of artificial intelligence into clinical practice and the development of robust and generalizable models.

Anichini et al. [15] investigated radiomics approaches for differentiating various focal liver lesions, including cysts and tumors, and discussed current diagnostic limitations and their therapeutic implications.

Early detection and characterization of hepatic lesions primarily rely on non-invasive imaging techniques, including ultrasound, CT, and MRI, while histopathological analysis remains the reference standard in cases of diagnostic uncertainty [13,16–19]. Despite advances in medical imaging, the differentiation between hepatic cysts and tumors remains challenging, particularly for small lesions or those with atypical or complex features, due to overlapping radiological signatures and the limited specificity of certain diagnostic criteria [20–24]. Multiphase contrast-enhanced CT is considered a reference modality for liver lesion evaluation due to its wide availability and high spatial resolution. Nevertheless, diagnostic ambiguities persist, particularly in cases such as hemorrhagic cysts, necrotic lesions, or suboptimal acquisitions, owing to overlapping imaging characteristics and variability in enhancement patterns [25,26].

Significant interobserver variability has been reported in the delineation and interpretation of hepatic lesions in imaging, representing an additional source of diagnostic uncertainty. This variability depends on factors such as radiologist experience, image

quality, and lesion complexity. In cases of uncertainty, this may lead to additional costly examinations (e.g., MRI, PET, repeated follow-up) or invasive procedures such as liver biopsy, which remains a reference standard but carries inherent risks [27–30].

Recently, machine learning and deep learning have been widely employed to develop robust solutions for the automated analysis of medical images without direct human intervention, by learning complex data representations. Convolutional neural networks (CNNs) are particularly effective for liver and lesion segmentation, tumor detection and classification, and the extraction of intricate features that are not discernible to the human eye. These approaches enable the modeling of complex patterns and improve diagnostic accuracy, reproducibility, and clinical decision support [31–35].

3. Materials and Methods

3.1. Patient Cohort and Data Collection

In this study, two groups of computed tomography (CT) images were analyzed using both prospective and retrospective approaches. The first group consisted of patients presenting with hepatic metastases, whereas the second group included patients diagnosed with pathological hepatic cysts. A summary of the characteristics, inclusion, and exclusion criteria of the two hepatic lesion datasets is shown in Table 1.

Table 1. Summary of the characteristics, inclusion, and exclusion criteria of the two hepatic lesion datasets.

Characteristics	Pathological Hepatic Cysts Dataset	Hepatic Metastases Dataset
Data source	University Hospital of Fez, Sidi Mohamed Ben Abdellah University	Ibn Sina University Hospital and National Institute of Oncology, Rabat
Inclusion criteria	Age > 18 years; underwent CT examinations in the radiology department of Hassan II University Hospital, Sidi Mohammed Ben Abdellah University; subsequently underwent surgery or histopathological biopsy to confirm the nature of hepatic lesions; confirmed diagnosis of pathological hepatic cysts based on clinical and radiological assessment; adequate quality liver imaging; complete lesion visualization	Age > 18 years; indication for hepatectomy due to hepatic or biliary tumors; availability of preoperative contrast-enhanced liver CT in DICOM format; provision of written informed consent; confirmed diagnosis of hepatic metastases based on oncological evaluation and radiological findings; available imaging data with visible metastatic lesions
Exclusion criteria	Low-quality images; incomplete examinations, unclear diagnosis; presence of confounding hepatic lesions	Incomplete data; prior preoperative treatment; presence of significant imaging artifacts; cases in which image reconstruction was deemed unsuitable; low-quality images; incomplete examinations; uncertain metastatic diagnosis
Lesion characteristics	Fluid-filled lesions with well-defined margins, size, and location within the liver parenchyma	Solid metastatic lesions with morphology, size, distribution, and enhancement patterns

3.2. Metastatic Liver Tumor Group

This group comprised 50 patients whose CT images were obtained from two distinct sources. First, 35 CT scans were selected from the publicly available Liver Tumor Segmentation Challenge (LiTS) dataset [36] and used for model training. Second, an internal dataset of 15 patients was collected as part of a prospective study conducted at the Department of Digestive Oncological Surgery (DCOD) of the National Institute of Oncology in Rabat, Morocco.

This internal dataset was derived from an initial cohort of 112 patients with chronic liver disease who underwent contrast-enhanced abdominal CT between January 2020 and July 2021. To ensure patient confidentiality, all data were fully anonymized. Data collection was carried out by a multidisciplinary team consisting of three specialist physicians and two computer scientists, in accordance with international ethical standards and the Declaration of Helsinki. Ethical approval was obtained from the Ethics Committee of the Faculty of

Medicine and Pharmacy of Rabat (Protocol No. 38/20), and the study adhered to the COREQ guidelines.

The inclusion criteria were as follows: age > 18 years, indication for hepatectomy due to hepatic or biliary tumors, availability of preoperative contrast-enhanced liver CT in DICOM format, and provision of written informed consent. Exclusion criteria included incomplete data, prior preoperative treatment, the presence of significant imaging artifacts, and cases in which image reconstruction was deemed unsuitable. Following this selection process, 15 patients were retained.

3.3. Pathological Hepatic Cyst Group

This group also included 50 patients, whose CT images were collected from an internal dataset as part of a retrospective study conducted at the Hassan II University Hospital in Fez, Morocco.

Data collection was performed by a multidisciplinary team composed of three specialist physicians and two computer scientists, in compliance with international ethical standards and the Declaration of Helsinki. In accordance with the recommendations of the ethics committee and the institutional review board, all data were anonymized by removing personal identifiers. Ethical approval was obtained from the Ethics Committee of the Faculty of Medicine and Pharmacy of Casablanca (Reference No. 17/15). Written informed consent was obtained from all patients.

The study included patients who underwent CT examinations in the radiology department of Hassan II University Hospital, Sidi Mohammed Ben Abdellah University, between January 2021 and July 2022, and who subsequently underwent surgery or histopathological biopsy to confirm the nature of hepatic lesions.

Among 521 medical records reviewed, 50 patients were selected according to predefined criteria. Inclusion criteria targeted patients identified with the following keyword: pathological hepatic cysts. Exclusion criteria included missing clinical data (age, sex, lesion location, lesion size, tumor stage, smoking status), prior preoperative treatment, poor quality CT images with significant artifacts, and patients who had received neoadjuvant chemotherapy. The patient cohort selection process is illustrated in Figure 1.

3.4. CT Acquisition Protocol and Image Preprocessing

CT acquisitions were performed using multi-detector scanners following a standardized protocol. Imaging parameters included a tube voltage of 120 kVp, automatic exposure control, and a slice thickness ranging from 1 to 1.25 mm. Acquisition settings were maintained as consistently as possible across examinations to minimize variability related to different scanners and imaging conditions.

To ensure data homogeneity and facilitate interpatient comparison, all CT images were resampled to an isotropic voxel size of $1 \times 1 \times 1 \text{ mm}^3$ using cubic interpolation. This preprocessing step was conducted in accordance with the recommendations of the Image Biomarker Standardization Initiative (IBSI) and reflects standard practices in radiomics to optimize the reproducibility and robustness of extracted features [37].

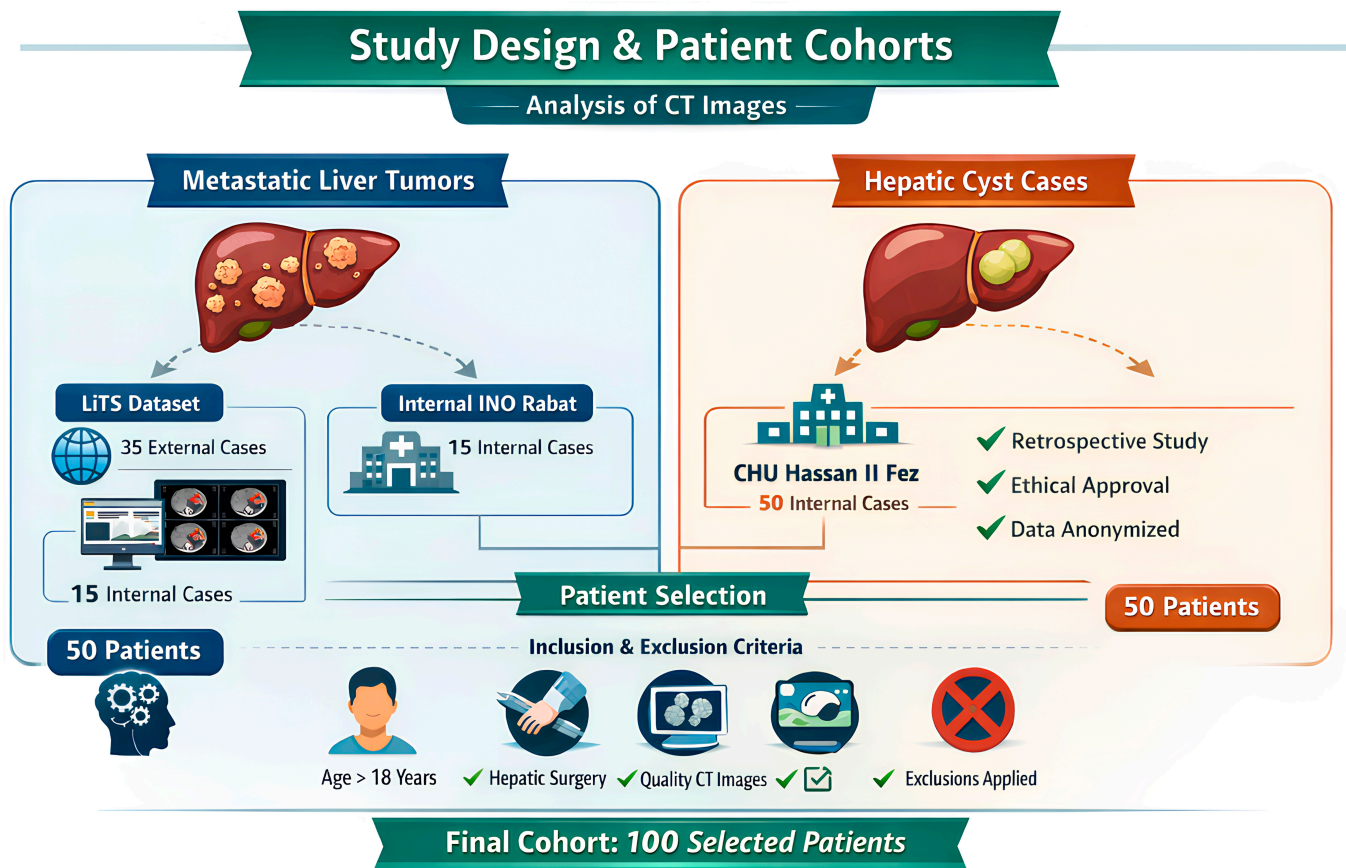


Figure 1. Study design and patient cohorts.

3.5. Proposed Model

To improve the predictive performance of the system, we propose an end-to-end convolutional neural network (CNN) based framework integrating deep transfer learning, model stacking, and decision fusion strategies.

3.6. Stacking Model and Deep Transfer Learning (Base Learners)

Initially, several convolutional neural networks and transformer-based architectures were selected as base learners, including ResNet-50, DenseNet-121, EfficientNet, VGG16/19, and Vision Transformer (ViT). These models were initialized using pretrained weights and subsequently fine-tuned on the considered medical imaging dataset. Each model learns distinct feature representations, promoting diversity in predictions, which is a key requirement for the effectiveness of ensemble methods.

The selection of these deep architectures is motivated by their demonstrated effectiveness in transfer learning applications within medical imaging, particularly in scenarios with limited annotated data. The combination of these architectures with transfer learning significantly enhances diagnostic accuracy, robustness, and generalization capability, as reported in recent studies [38–40]. Indeed, these models, originally pretrained on large-scale natural image datasets such as ImageNet, enable the transfer of generic visual representations to domain-specific medical tasks, thereby reducing dependence on costly and limited annotated datasets.

ResNet and DenseNet architectures leverage residual and dense connections, respectively, facilitating gradient propagation and improving the learning of discriminative deep features, even in low data regimes. These mechanisms effectively mitigate the vanishing gradient problem and promote feature reuse [41–43]. In contrast, VGG architectures, despite their relative simplicity, rely on a homogeneous structure based on stacked convo-

lutional layers, enabling efficient hierarchical feature extraction from low-level patterns to high-level semantic representations [44]. EfficientNet introduces a compound scaling strategy that simultaneously optimizes network depth, width, and resolution, resulting in an excellent trade-off between performance and computational efficiency [45–47]. More recently, Vision Transformers have demonstrated their ability to capture global dependencies in images through attention mechanisms, allowing effective modeling of long-range relationships between different regions and improving the analysis of complex structures in medical imaging [48–50].

3.7. Decision Fusion and Stacking Strategy

To enhance the robustness and overall performance of the classification system, a hybrid strategy combining decision fusion and stacking was adopted. These ensemble approaches leverage the complementarity of multiple deep models, mitigating individual errors while improving generalization capability.

3.7.1. Decision Fusion

Decision fusion, including strategies such as majority voting, weighted averaging, and stacking, enables the combination of predictions from multiple models to reduce individual errors and improve overall robustness. These ensemble techniques exploit model complementarity to provide more reliable and accurate decisions, generally outperforming individual models [51–54].

Formally, let M denote the number of models and $p_i(x)$ the probability predicted by the i th model for an input x . Several fusion rules were explored:

- Majority Voting: The final class corresponds to the class predicted by the majority of models.
- Soft Voting (Probability Averaging): The final prediction is obtained by averaging the predicted probabilities across models.

$$\hat{y} = \arg \max_c \left(\frac{1}{M} \left(\sum_{i=1}^M p_i^c(x) \right) \right) \quad (1)$$

- Weighted Average: Each model is weighted according to its individual performance:

$$\hat{y} = \operatorname{argmax}_c \left(\sum_{i=1}^M w_i \cdot p_i^c(x) \right) \quad (2)$$

where w_i denotes the weight assigned to the i th model, determined based on its validation performance.

These strategies contribute to reducing the variance of the overall model and improving prediction stability, particularly in settings characterized by limited and noisy data.

To further enhance robustness, a weighted majority voting scheme was implemented, in which each model contributes according to a weight proportional to its performance during cross-validation. Consequently, the most accurate models exert a greater influence on the final decision, thereby mitigating the impact of less reliable models. This adaptive weighting strategy, illustrated in Figure 2, improves overall classification accuracy compared to standard majority voting. The final fusion rule corresponds to majority voting across the five primary networks. The predicted class is determined based on the highest number of votes. In the case of ties (<0.5%), the decision is resolved by selecting the class with the highest average softmax probability among the tied models.

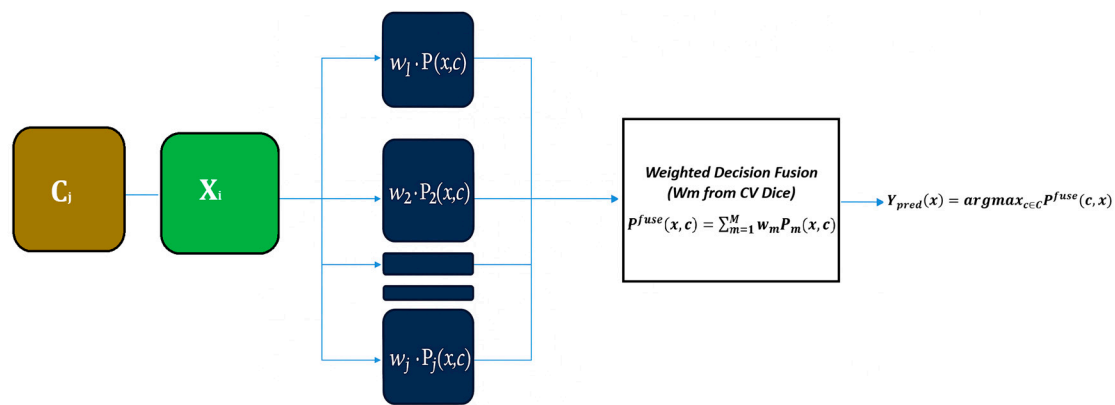


Figure 2. Weighted decision merging workflow.

Using the majority voting method, the exit decision $D_{i,j}$ is given by Equation (3):

$$D_{i,j} = \begin{cases} 1 & \text{if } p\left(\frac{\omega_n^*}{x_i}\right) = \max_{j=1}^n p\left(\frac{\omega_j}{x_i}\right) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

In this case, the weight ω_j will be associated with the shape Z if:

$$w_j = \max_{j \in \{1, \dots, n\}} \sum_{i=1}^R \delta(d_i, \omega_j) \quad (4)$$

where $\delta(d_i, \omega_j)$ is an indicator function taking the value 1 when the classifier's decision corresponds to the class ω_j and 0 otherwise. Thus, the class assigned to form Z corresponds to the one receiving the most votes among the R classifiers with:

$$\sum_{i=1}^R D_{i,j} = \max_{n=1}^j \sum_{i=1}^R D_{i,n} \quad (5)$$

The right-hand side of Equation (5) corresponds to the sum of the votes of each network composing the stacking model, thus allowing the model to make a final decision based on the majority vote.

3.7.2. Model Stacking

Stacking is a more advanced approach to model fusion, in which the predictions of the base models are used as inputs to a meta model (meta learner). More formally, for a set of predictions $\{p_1(x), p_2(x), \dots, p_M(x)\}$, a second model g is trained to learn the optimal combination function:

3.8. Overall Pipeline of the Proposed Model

The proposed pipeline consists of the following steps, as shown in Figure 3:

- Pre-processing and normalization of the input medical image;
- Training of the base models with transfer learning;
- Generation of individual predictions on the validation set;
- Application of weighted majority voting;
- Performance evaluation on an independent test set.

The ResNet 50, DenseNet-121, EfficientNet, VGG-16/19, and ViT architectures are trained directly on medical images. Each model produces an independent prediction, without explicit feature extraction. The exact method applied is weighted majority voting, in which each architecture contributes with a weight proportional to its cross-validation

performance. Stronger models thus have a greater influence on the final decision, reducing the effect of weaker models. This adaptive weighting, illustrated in Figure 3, improved the overall accuracy of the classification compared to simple majority voting.

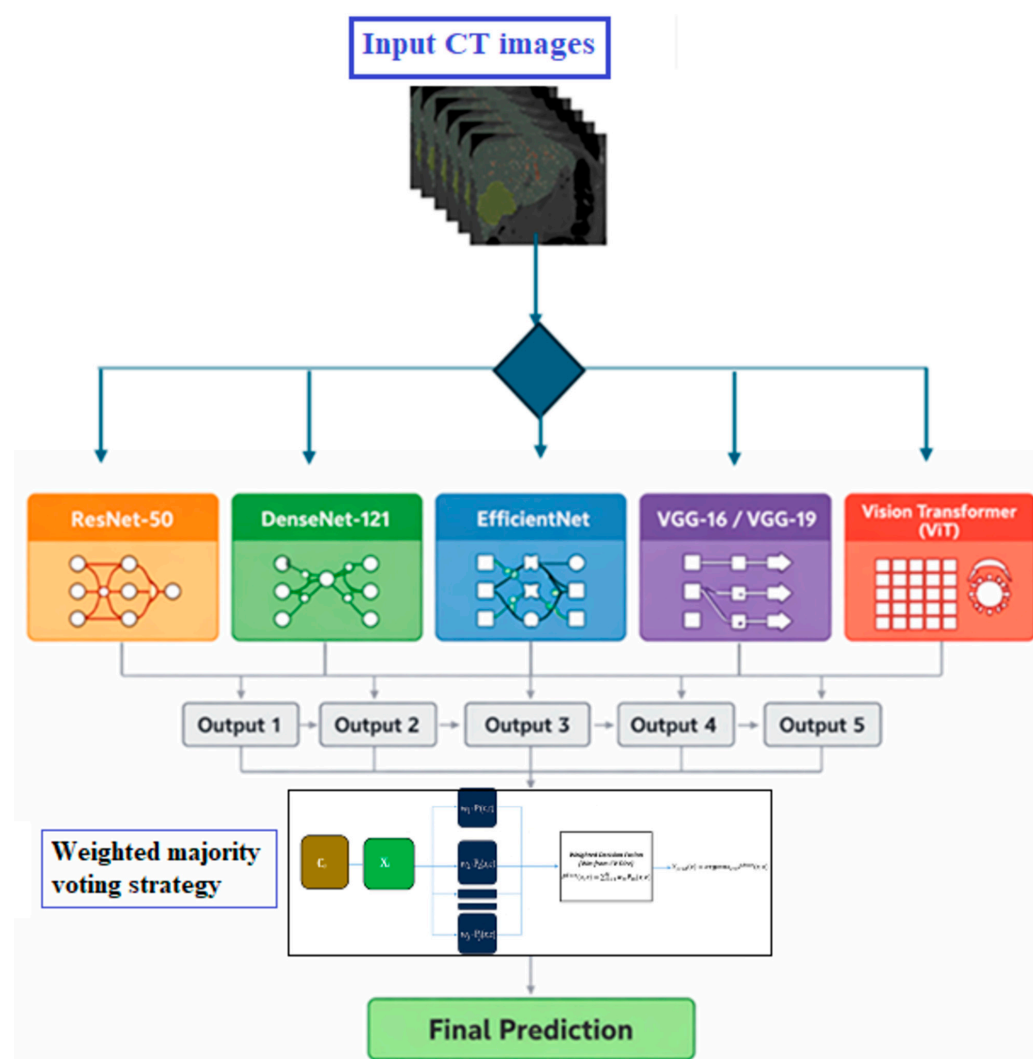


Figure 3. Workflow of the proposed method.

3.9. Precision Diffusion Classification Model Workflow with Deep Model Stacking

We improved the classification performance of our model by combining model stacking, deep transfer learning, and precision fusion. This approach selects the prediction with the most votes from the five predictions in the model set as the final result. The workflow of our model, illustrated in Figure 4, comprises two key processes: training and testing. During training, we used 5-fold cross-validation on a set of 16,000 slices from 80 CT images. During the testing process, we evaluated the model on the 20 remaining undamaged CT images.

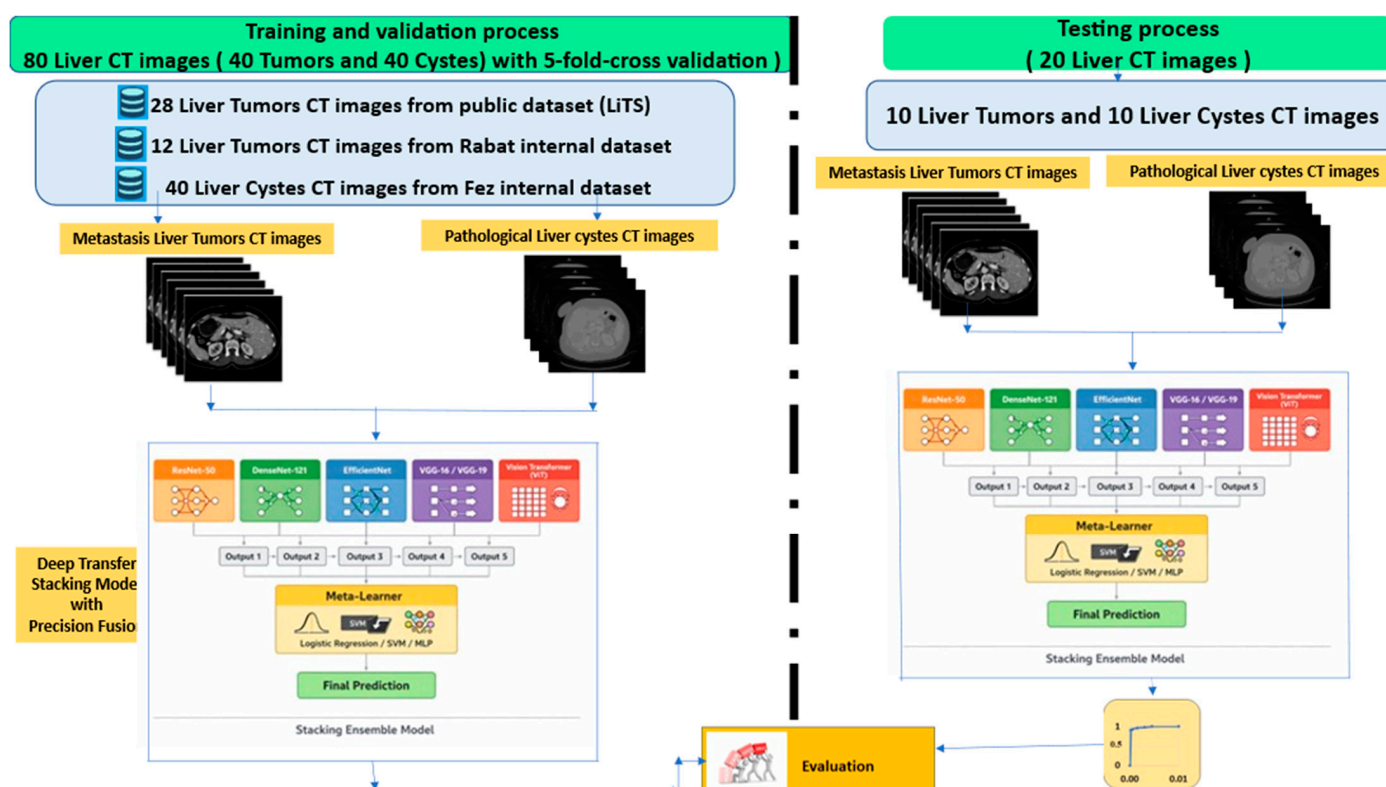


Figure 4. Workflow of the precision diffusion classification model with deep model stacking.

4. Results

4.1. Distribution of Training and Test Datasets

The 100 patients included in the reference dataset were randomly divided into two independent subsets: an 80% training set and a 20% testing set, as summarized in Table 2. Importantly, the dataset partitioning was performed at the patient level rather than at the image slice level to prevent data leakage. Consequently, all CT slices belonging to a given patient were exclusively assigned to either the training or testing set, ensuring that no images from the same patient were present in both subsets. This strategy provides a more reliable evaluation of the proposed model performance and avoids potential overestimation of classification accuracy.

Table 2. Composition of the training and test sets.

Training Set Using 5-Fold Cross-Validation	Test Set
80 patients (80%)	20 patients (20%)
40 patients with metastatic liver tumors generating 6340 CT image slices	10 patients with metastatic liver tumors generating 1800 CT image slices
40 patients with hepatic pathological cysts generating 6460 CT image slices	10 patients with hepatic pathological cysts generating 1400 CT image slices
Total: 12,800 CT image slices	Total: 3200 CT image slices

4.1.1. Training Set

The training set consisted of 80 patients, 40 with confirmed metastatic liver tumors and 40 with pathological liver cysts. A total of 12,800 slices from the 80 CT images were selected to train the proposed model using 5-fold cross-validation.

4.1.2. Test Set

The test set consisted of 20 patients: 10 with confirmed metastatic liver tumors and 10 with pathological liver cysts. A total of 3200 slices from the 20 CT images were used to test the proposed model.

4.2. Experimental Testbed

The software developed for this study was implemented in Python 3.6 using TensorFlow r1.9 (Apache 2.0 license). The neural network training was performed on a workstation optimized for GPU computing, equipped with four NVIDIA GeForce GTX 1080 Ti cards (11 GB memory, Pascal microarchitecture; NVIDIA, Santa Clara, CA, USA). Inference speed performance was evaluated using a single GPU.

4.2.1. Training Phase Performance

Mean Binary Cross-Entropy Loss Function

To quantify the difference between the model predictions and the actual values, we will use the mean binary cross entropy L , widely used in binary classification tasks [55], as the loss function for model training. This function is defined as follows:

$$L = -\frac{1}{N} \sum_{i=1}^N [Y_i \log(\hat{Y}_i) + (1 - Y_i) \log(1 - \hat{Y}_i)] \quad (6)$$

where Y_i represents the actual label and $\hat{Y}_i$ the probability predicted by the model. The learning rate is initially set to 0.001, then decreased after four epochs if there is no improvement in the loss function, through the use of callback mechanisms. The models were parameterized to be trained over a total of 60 iterations.

To validate the training process of the proposed model, a five-fold cross-validation was implemented. The dataset was partitioned into five subsets of equal size. At each iteration, four folds were used for model training, while the remaining fold was used for validation. This process was repeated five times, with each fold used only once as the validation set. The final model performance was obtained by averaging the results from the five iterations, thus providing a robust and reliable assessment of the model's generalization capacity.

The average binary cross-entropy loss functions obtained during the training and validation processes for the different CNN architectures are shown in Figure 5 and Figure 6, respectively.

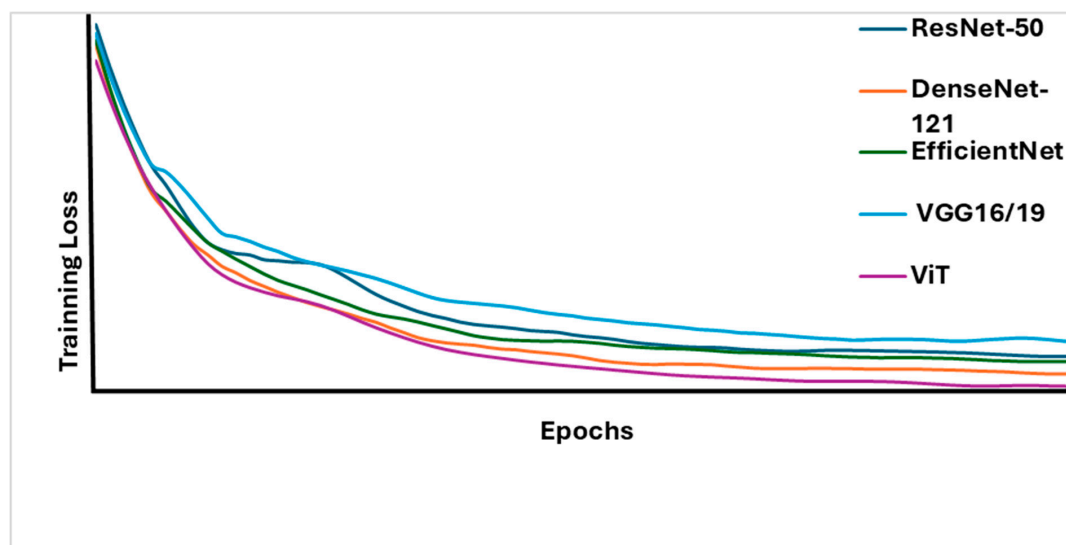


Figure 5. Average binary cross-entropy loss functions obtained during the training process.

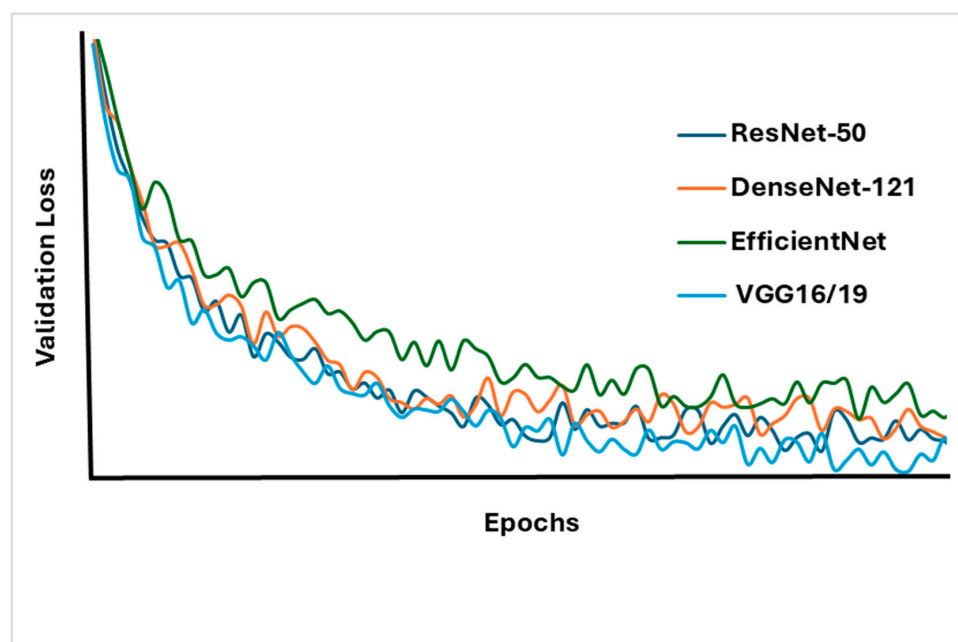


Figure 6. Average binary cross-entropy loss functions obtained during the validation process.

The average accuracy functions obtained during the training and validation processes for the different CNN models are shown in Figures 5 and 6, respectively.

4.2.2. Performance of the Proposed Model Test Phase

Confusion Matrices

The confusion matrix is a fundamental tool for evaluating the performance of binary classification models, particularly during the testing phase of multiple CNN models.

The performance of the five CNN models (ResNet-50, DenseNet-121, EfficientNet, VGG16/19, and ViT) compared to that of the proposed Stacking model was rigorously evaluated across the test set using the confusion matrix, allowing for precise quantification of correct and incorrect predictions, as shown in Figure 7.

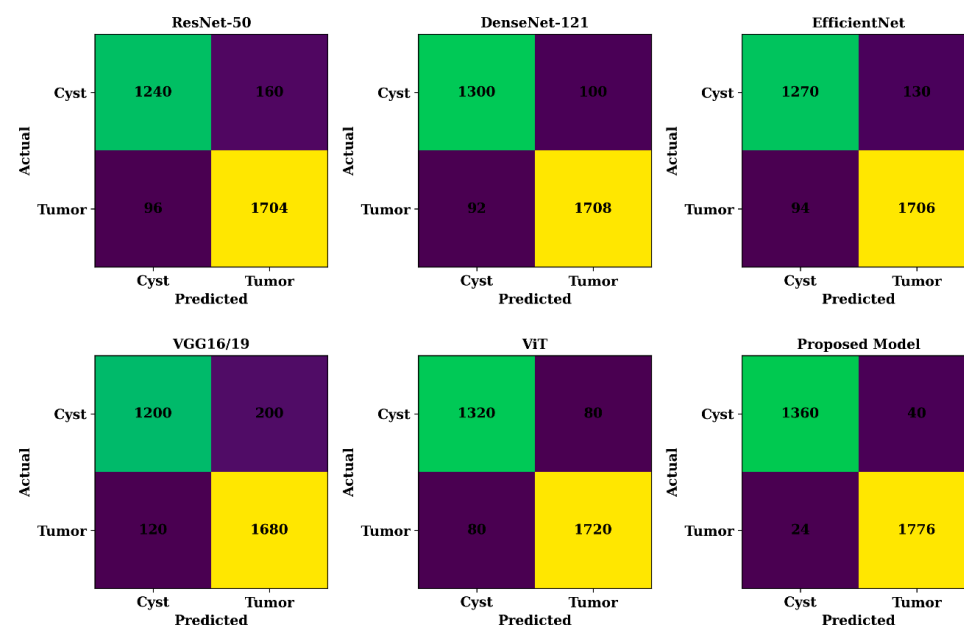


Figure 7. Confusion matrix of the six models: ResNet-50, DenseNet-121, EfficientNet, VGG16/19, ViT, and the proposed Stacking model.

Confusion matrices allow for a detailed analysis of the performance of different models by distinguishing between correct and incorrect classifications for the two classes: cyst (Cyst) and tumor (Tumor). A simple reading of these matrices shows that for:

- Classic CNN models:
 - ResNet-50: The model has a good tumor detection rate (TP = 1704), but a relatively high number of false positives (FP = 160) and false negatives (FN = 96). This indicates good sensitivity, but room for improvement in accuracy.
 - DenseNet-121: An improvement is observed with a reduction in errors (FP = 100, FN = 92). The model shows a better balance between tumor detection and the reduction in false classifications.
- EfficientNet: Performance is comparable to DenseNet-121, with a slight compromise (FP = 130, FN = 94). The model remains generally robust but slightly less accurate.
- VGG16/19:
 - This model exhibits the lowest performance, with the highest number of errors (FP = 200, FN = 120). This confirms its limited ability to capture complex features in medical images.
 - Vision Transformer (ViT) model: This model shows a marked improvement with a significant reduction in errors (FP = 80, FN = 80). It offers a good balance between sensitivity and specificity, demonstrating the effectiveness of global attention mechanisms.
- The proposed stacking model:

This model achieves the best overall performance: True Negatives (TN) = 1360, True Positives (TP) = 1776, False Positives (FP) = 40, False Negatives (FN) = 24. This reflects excellent tumor detection capabilities (very low FN), a significant reduction in false alarms (minimal FP), and overall performance superior to the other models.

4.3. Overall Comparative Analysis

We observe that the number of false negatives (FN) decreases progressively from VGG → ResNet → DenseNet → ViT → Stacking of the proposed model. This shows that our proposed model clearly reduced the number of missed tumor cases. Similarly, the number of false positives also decreases progressively from VGG → ResNet → DenseNet → ViT → Stacking of the proposed model, thus demonstrating that our proposed model reduced the number of erroneous diagnoses. Furthermore, the performance of our proposed model clearly outperforms the other models, with a 70% reduction in false negatives compared to VGG and an 80% reduction in false negatives compared to VGG.

The confusion matrices presented allow for a detailed analysis of the performance of the different models for the classification of cysts and tumors. Classical CNN models, such as ResNet-50, DenseNet-121, and EfficientNet, exhibit generally satisfactory performance, although they still produce a significant number of false positives and false negatives. In particular, ResNet-50 demonstrates good tumor detection capabilities, but with a relatively higher error rate, while DenseNet-121 and EfficientNet offer a better balance between sensitivity and specificity thanks to reduced classification errors. The VGG16/19 model exhibits the weakest performance, characterized by the highest number of false positives and false negatives, reflecting its limitations in extracting complex features. The Transformer-based model (ViT) significantly improves these results, with a notable reduction in errors, indicating a better ability to capture global dependencies in images. However, the proposed model outperforms all the compared approaches, registering the lowest number of false positives (40) and false negatives (24), as well as the highest values for true positives and true negatives. This performance indicates excellent class discrimination capabilities, with

a substantial reduction in critical errors, particularly false negatives, which are especially important in a clinical setting. These results confirm the superiority of the proposed model in terms of robustness, generalizability, and diagnostic reliability.

4.4. Ablation Study

To study the contribution of each component of the proposed framework, several ablation experiments were conducted. The effects of transfer learning and multi-CNN feature representation were evaluated by progressively removing each component and comparing the resulting performance with that of the complete model.

The objective of this analysis is to demonstrate the effectiveness of the proposed architecture compared with conventional individual CNN-based approaches and the ViT method. The ablation study provides experimental evidence of the contribution of each methodological component to the improvement of the final model performance, evaluated using accuracy, precision, recall, F1-score, and AUC metrics.

Three experiments were conducted. The first experiment (baseline CNN) consisted of simplifying the proposed framework by retaining only a single CNN architecture from the complete model in order to evaluate the contribution of the individual network performance. The second experiment (without transfer learning) aimed to assess the impact of transfer learning by training a single CNN architecture from scratch, without pretrained weights, and analyzing its performance. Finally, the third experiment (complete proposed framework) corresponded to the proposed model, integrating pretrained CNN architectures, transfer learning, and the decision fusion strategy. The comparison between these configurations, presented in Table 3, allowed us to validate the contribution of the proposed components and demonstrate the robustness of the final model.

Table 3. Performance of the ablation study experiments.

Experiment	Model Configuration	Performances			
		Accuracy	Precision	Recall	Score F ₁
Baseline CNN	ResNet-50	0.92	0.91	0.95	0.93
	DenseNet-121	0.94	0.94	0.95	0.95
	EfficientNet	0.93	0.93	0.95	0.94
	VGG16/19	0.90	0.89	0.93	0.91
	ViT	0.95	0.96	0.96	0.96
Without transfer learning	ResNet-50	Could not be achieved due to training instability after 80 training iterations			
	DenseNet-121				
	EfficientNet				
	VGG16/19				
	ViT				
Complete proposed model	Complete proposed model	0.98	0.98	0.99	0.98

Table 3 highlights the comparative performance of this study's ablation configurations. In the Baseline CNN experiment overall, all models demonstrate high performance, with accuracy values greater than 0.90, indicating good discrimination between classes.

Among the classic architectures, DenseNet-121 and EfficientNet offer robust performance, with respective F1 scores of 0.95 and 0.94, reflecting a good balance between precision and sensitivity. ResNet-50 also shows high sensitivity (0.95), indicating good tumor detection, although its slightly lower precision (0.91) suggests a higher number of false positives. Conversely, VGG16/19 exhibits the lowest performance among the CNN models, with an accuracy of 0.90 and an F1 score of 0.91, which can be explained by a less

effective representation capacity when faced with complex data. The ViT (Vision Transformer) model outperforms traditional CNN architectures, with an accuracy of 0.95 and an F1 score of 0.96, demonstrating the effectiveness of attentional mechanisms in capturing global dependencies in medical images.

In the second experiment, reliable performance evaluation of this configuration could not be achieved due to training instability, as the network required a significantly higher number of iterations than the predefined 80 training epochs to reach convergence. This behavior highlights the difficulty of training deep CNN models from scratch when dealing with limited medical imaging datasets.

Finally, in the third experiment, the proposed model achieves the best overall performance, with an accuracy of 0.98, a precision of 0.98, a sensitivity of 0.99, and an F1 score of 0.98. These results indicate not only excellent classification capabilities but also a significant reduction in false positives and false negatives. The particularly high sensitivity (0.99) is crucial in a medical context, as it reflects a near-optimal ability to classify liver diseases, thus minimizing the risk of missed diagnoses. Compared to the best competing model (ViT), the complete proposed model represents a gain of:

- +3% in accuracy (0.98 vs. 0.95);
- +2% in F1 score (0.98 vs. 0.96);
- +3% in precision (0.98 vs. 0.96).

4.5. Comparative Performance Analysis (Gains in %)

To quantify the improvements brought about by the proposed model, an analysis of relative gains (%) was performed compared to the reference models. The relative gain is defined as follows:

$$Gain (\%) = \frac{M_{proposed\ model} - M_{reference}}{M_{baseline}} \times 100 \quad (7)$$

where $M_{reference}$ represents the metric considered (Accuracy, Precision, Recall, or F1-score).

The relative gains (%) of the proposed model compared to the reference models are given in Table 4.

Table 4. Relative gains (%) of the proposed model compared to the reference models.

	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
ResNet-50	+6.52	+7.69	+4.21	+5.38
DenseNet-121	+4.26	+4.26	+4.21	+3.16
EfficientNet	+5.38	+5.38	+4.21	+4.26
VGG16/19	+8.89	+10.11	+6.45	+7.69
ViT	+3.16	+2.08	+3.12	+2.08

The analysis of relative gains highlights a systematic improvement of the proposed model compared to all reference architectures. The most significant gains are observed compared to VGG16/19, confirming the limitations of classical deep architectures without advanced feature extraction mechanisms.

Furthermore, although the gains compared to ViT are more moderate, they remain significant in a context where performance is already high, underscoring the robustness and generalizability of the proposed model.

In particular, the improved sensitivity (+3% to +4%) is crucial in a medical setting, as it reduces the risk of false negatives. Similarly, the increased accuracy indicates a decrease in false positives, thus improving the system's clinical reliability.

4.6. Statistical Analysis of Performance

To assess the robustness of the observed performance, 95% confidence intervals (95% CIs) were estimated for the accuracy of each model. The confidence interval is defined as:

$$IC_{95\%} = a \pm 1.96 \sqrt{\frac{a(1-a)}{N}} \quad (8)$$

where a represents the accuracy and N the total number of samples.

Furthermore, a statistical comparison between the models was performed using McNemar's test (Table 5), adapted for comparing classifiers on the same test set:

$$\chi^2 = \frac{(|b - c| - 1)^2}{b + c} \quad (9)$$

where b and c represent, respectively, the number of samples misclassified by one model but correctly classified by the other, and vice versa.

Table 5. Statistical analysis of model performance.

	Accuracy	IC 95%	p-Value
ResNet-50	0.92	[0.910–0.930]	<0.001
DenseNet-121	0.94	[0.932–0.948]	<0.01
EfficientNet	0.93	[0.922–0.938]	<0.01
VGG16/19	0.90	[0.888–0.912]	<0.001
ViT	0.95	[0.942–0.958]	<0.05
Modèle proposé	0.98	[0.975–0.985]	—

The narrow confidence intervals observed for all models indicate good stability of the estimates. The proposed model not only exhibits the best performance in terms of accuracy, but also a confidence interval that does not significantly overlap with those of the lower-performing models, suggesting a robust improvement.

The results of the McNemar test show that the differences between the proposed model and the other models are statistically significant ($p < 0.05$), particularly compared to classical CNN architectures. Even though the difference with the ViT model is smaller, it remains significant, confirming the superiority of the proposed model.

4.7. Area Under the ROC Curve (AUC-ROC)

The area under the ROC curve (AUC-ROC) is an overall indicator of the model's ability to distinguish between the two classes. Unlike metrics dependent on a single decision threshold, AUC-ROC allows us to assess the quality of the separation between positive and negative classes across all possible thresholds.

The ROC curve is based on two fundamental quantities, namely sensitivity and false positive rate. The area under the ROC curve, denoted AUC, is expressed as follows:

$$AUC = \int_0^1 TPR(x) dx \quad (10)$$

where TPR denotes the True Positive Rate, equivalent to the sensitivity, and x represents the False Positive Rate.

From a probabilistic perspective, the AUC can also be interpreted as the probability that a randomly chosen positive sample will receive a higher prediction score than a randomly chosen negative sample.

A high AUC value reflects a better ability of the model to distinguish between the two classes.

In this study, the AUC-ROC was calculated (Figure 8) for each of the five models (ResNet-50, DenseNet-121, EfficientNet, VGG16/19, and ViT) and for the proposed Stacking model to provide a synthetic and robust measure of their classification performance across the test set.

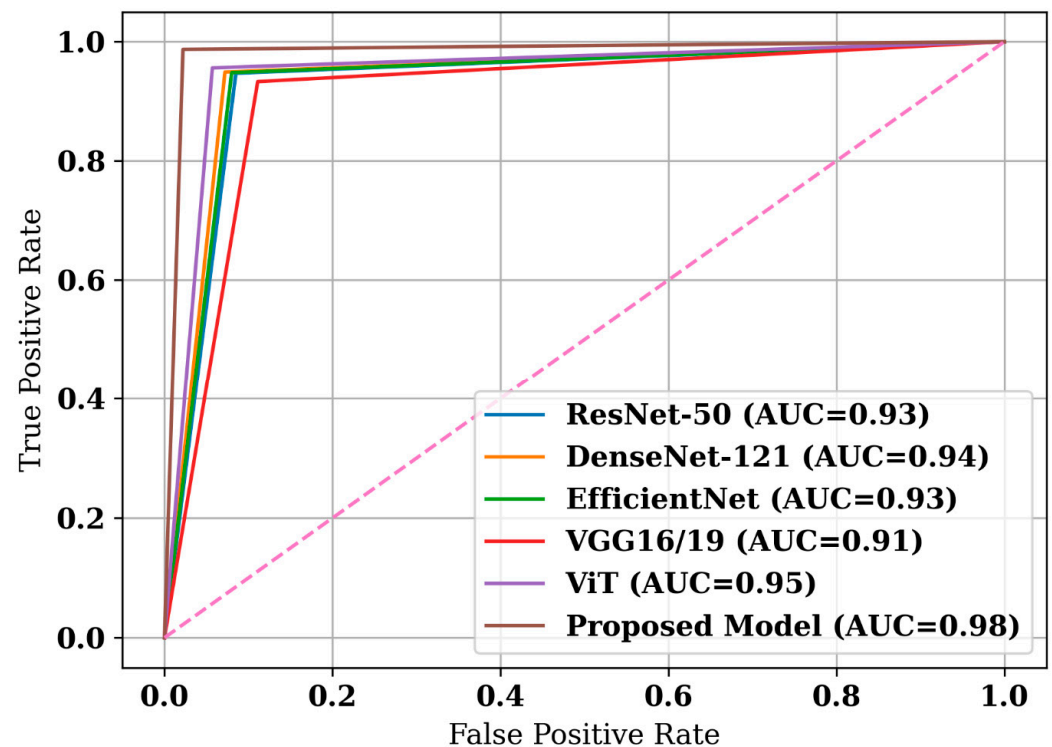


Figure 8. Calculated AUC-ROC for each of the models: ResNet-50, DenseNet-121, EfficientNet, VGG16/19, and ViT, and the proposed stacking model.

Among the classic CNN architectures, DenseNet-121 achieves the best performance (AUC = 0.94), followed by ResNet-50 and EfficientNet (AUC = 0.93), reflecting a good overall ability to discriminate between cyst and tumor classes. In contrast, VGG16/19 has the lowest value (AUC = 0.91), suggesting lower efficiency in modeling complex discriminating features.

The transformer-based model, ViT, further improves performance with an AUC of 0.95, highlighting the value of global attention mechanisms for capturing complex spatial relationships in medical images.

The proposed stacking model achieves the best overall performance with an AUC of 0.98, indicating near-optimal discrimination between the two classes of cyst and tumor. This significant improvement compared to other models suggests that combining predictions from multiple architectures leverages their complementarities and enhances the overall robustness of the system.

Analysis of relative gains in AUC-ROC highlights a systematic improvement of the proposed stacking model compared to all individual architectures. The most significant gains are observed compared to VGG16/19 (+7.69%), confirming the limitations of classical CNN architectures in modeling complex relationships.

Even compared to the highest-performing models, such as ViT, the proposed model maintains a notable advantage (+3.16%), which is particularly significant in a context where performance is already high. Indeed, marginal improvements at this AUC level generally translate into substantial gains in clinical discrimination capability.

These results demonstrate that the stacking approach leverages the complementarity of the different models, thus improving the robustness and generalizability of the system.

An AUC of 0.98 reflects near-optimal performance, reinforcing the potential of the proposed model for reliable application in computer-assisted diagnostics.

5. Discussion

The results obtained in this study clearly demonstrate the effectiveness of deep learning approaches for classifying liver lesions from CT images in low-resource areas. All the evaluated models achieved high performance, with accuracy values greater than 0.90 and AUC-ROC values greater than 0.91, confirming their ability to extract relevant discriminating features.

Among the classic CNN architectures, DenseNet-121 and EfficientNet showed robust performances, which can be attributed to their information propagation optimization mechanisms and their ability to effectively exploit multi-scale features. Conversely, VGG16/19 exhibited the lowest performance, likely due to its older architecture and its lesser suitability for the complexity of medical data. ResNet-50, while performing well, showed a slight tendency to generate more false positives, which may affect overall accuracy.

The transformer-based model (ViT) outperformed traditional CNN architectures, highlighting the value of global attention mechanisms for capturing complex spatial relationships in medical images. This observation is consistent with recent work emphasizing the superiority of transformers in certain medical vision tasks.

However, the proposed stacking model achieved the best overall performance (Accuracy = 0.98, F1-score = 0.98, AUC = 0.98), demonstrating a significant improvement over all individual models. This superiority can be explained by stacking's ability to combine the complementary strengths of the different models, thus reducing architecture-specific errors. In particular, the notable reduction in false negatives observed in the confusion matrix is of major importance in a clinical setting, where the failure to detect a tumor can have serious consequences.

Analysis of relative gains (%) and statistical metrics confirmed that the improvements provided by the proposed model are consistent and significant. Furthermore, the high AUC-ROC values indicate excellent discrimination capabilities independent of the decision threshold, enhancing the model's robustness.

However, some limitations must be considered. First, the dataset size remains relatively small, which may affect the generalizability of the results. Additionally, the study relies solely on 2D CT images, without incorporating clinical or multimodal information that could further improve performance. Finally, while the stacking model offers superior performance, it may exhibit higher computational complexity, potentially limiting its real-time deployment.

Despite these limitations, the results are promising and suggest that the proposed approach could provide a reliable solution for rapid, non-invasive, and cost-effective diagnosis of liver lesions. Future work should include larger, multicenter databases and the integration of clinical and radiomics data to further improve the model's robustness and clinical relevance.

6. Conclusions

In this study, we proposed an approach based on a stacking model combining several deep learning architectures, including convolutional neural networks (ResNet-50, DenseNet-121, EfficientNet, VGG16/19) and a transformer-based model (ViT), to improve the classification of liver lesions into cysts and tumors from CT images.

The experimental results demonstrated that, although the individual models exhibit high performance, the proposed model consistently outperforms all the approaches compared. In particular, it achieves an accuracy of 0.98, an F1 score of 0.98, and an AUC-ROC of

0.98, reflecting near-optimal discrimination capabilities. Analysis of the confusion matrices also revealed a significant reduction in false positives and false negatives, which is crucial in a clinical setting. Statistical evaluation, including confidence intervals and significance tests, confirmed that the observed improvements are robust and not due to chance. These results highlight the value of the stacking approach, which leverages the complementarity of different models to improve the generalization and reliability of predictions. Thus, the proposed model represents a promising solution for rapidly differentiating pathological liver cysts from metastatic liver tumors in a non-invasive and cost-effective manner, with strong potential for integration into clinical systems in low-resource areas.

Author Contributions: Conceptualization: A.B. (Amine Benkabbou), M.Q., R.M. and A.C.; Methodology, A.B. (Amine Benkabbou), M.Q. and A.C.; Software, A.B. (Anass Benfares); Validation: A.M., Z.E.M., M.A.E.A.E.H., O.L., A.B. (Amine Benkabbou), M.Q., A.S. and A.C., formal analysis, A.B. (Amine Benkabbou), M.Q., M.O., O.J.M. and A.L.; investigation, A.B. (Amine Benkabbou), R.M. and M.Q.; writing—review and editing: A.L., H.Q. and O.J.M.; supervision: A.B. (Amine Benkabbou), M.Q. and A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Research Ethics Committee of the Faculty of Medicine and Pharmacy of Casablanca (code: 17/15 and date of approval: 28 April 2015) and Ethical approval was obtained from the Ethics Committee of the Faculty of Medicine and Pharmacy of Rabat (Protocol No. 38/20 and date of approval: 21 May 2026).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available upon reasonable request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bray, F.; Laversanne, M.; Sung, H.; Ferlay, J.; Siegel, R.L.; Soerjomataram, I.; Jemal, A. Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A Cancer J. Clin.* **2024**, *74*, 229–263. [[CrossRef](#)] [[PubMed](#)]
2. Pfahl, E.L.; Pracha, N.S.; Emlendi, M.H.; Le, P.-H.D.; Makary, M.S. Artificial Intelligence in the Diagnosis and Prognostic Stratification of Hepatocellular Carcinoma: Current Evidence, Clinical Applications, and Future Perspectives. *Biomedicines* **2026**, *14*, 505. [[CrossRef](#)] [[PubMed](#)]
3. Lee, J.M.; Bae, J.S. Enhancing Diagnostic Precision in Liver Lesion Analysis Using a Deep Learning-Based System: Opportunities and Challenges. *Nat. Rev. Clin. Oncol.* **2024**, *21*, 485–486. [[CrossRef](#)] [[PubMed](#)]
4. Saba, L.; Biswas, M.; Kuppili, V.; Cuadrado Godia, E.; Suri, H.S.; Edla, D.R.; Omerzu, T.; Laird, J.R.; Khanna, N.N.; Mavrogeni, S.; et al. The Present and Future of Deep Learning in Radiology. *Eur. J. Radiol.* **2019**, *114*, 14–24. [[CrossRef](#)] [[PubMed](#)]
5. Aggarwal, R.; Sounderajah, V.; Martin, G.; Ting, D.S.W.; Karthikesalingam, A.; King, D.; Ashrafian, H.; Darzi, A. Diagnostic Accuracy of Deep Learning in Medical Imaging: A Systematic Review and Meta-Analysis. *npj Digit. Med.* **2021**, *4*, 65. [[CrossRef](#)] [[PubMed](#)]
6. Wang, K.; Lu, X.; Zhou, H.; Gao, Y.; Zheng, J.; Tong, M.; Wu, C.; Liu, C.; Huang, L.; Jiang, T.; et al. Deep Learning Radiomics of Shear Wave Elastography Significantly Improved Diagnostic Performance for Assessing Liver Fibrosis in Chronic Hepatitis B: A Prospective Multicentre Study. *Gut* **2019**, *68*, 729–741. [[CrossRef](#)] [[PubMed](#)]
7. Huang, J.; Xiang, Y.; Gan, S.; Wu, L.; Yan, J.; Ye, D.; Zhang, J. Application of artificial intelligence in medical imaging for tumor diagnosis and treatment: A comprehensive approach. *Discov. Oncol.* **2025**, *16*, 1625. [[CrossRef](#)] [[PubMed](#)]
8. Qjidaa, M.; Benfares, A.; El Azami El Hassani, M.A.; Benkabbou, A.; Souadka, A.; Majbar, A.; El Moatassim, Z.; Oumlaz, M.; Lahnaoui, O.; Mouhcine, R.; et al. High-Precision, Automatic, and Fast Segmentation Method of Hepatic Vessels and Liver Tumors from CT Images Using a Fusion Decision-Based Stacking Deep Learning Model. *BioMedInformatics* **2025**, *5*, 53. [[CrossRef](#)]
9. Rizzo, S.; Botta, F.; Raimondi, S.; Origgi, D.; Fanciullo, C.; Morganti, A.G.; Bellomi, M. Radiomics: The Facts and the Challenges of Image Analysis. *Eur. Radiol. Exp.* **2018**, *2*, 36. [[CrossRef](#)] [[PubMed](#)]

10. Chenin, M.; Paisant, A.; Lebigot, J.; Bazeries, P.; Debbi, K.; Ronot, M.; Laurent, V.; Aubé, C. Cystic Liver Lesions: A Pictorial Review. *Insights Imaging* **2022**, *13*, 116. [CrossRef] [PubMed]
11. Anderson, M.A.; Dhami, R.S.; Fadzen, C.M.; Molina, G.; Taylor, M.S.; Deshpande, V.; Qadan, M.; Catalano, O.A.; Ferrone, C.R.; Mojtahed, A. CT and MRI features differentiating mucinous cystic neoplasms of the liver from pathologically simple cysts. *Clin. Imaging* **2021**, *76*, 46–52. [CrossRef] [PubMed]
12. Wang, L.; Xu, Z.; Zhang, L.; Zhao, K.; Sun, H.; Pan, Z.; Li, Q. Differentiation of Malignant from Benign Focal Liver Lesions in Triphase-Enhanced CT Using Machine-Learning-Based Radiomics. *Br. J. Radiol.* **2025**, *98*, 1623–1631. [CrossRef] [PubMed]
13. Freitas, P.S.; Janicas, C.; Veiga, J.; Matos, A.P.; Herédia, V.; Ramalho, M. Imaging evaluation of the liver in oncology patients: A comparison of techniques. *World J. Hepatol.* **2021**, *13*, 1936–1955. [CrossRef] [PubMed]
14. Alshomrani, F.; Alghamdi, A.; Alsubhi, M.; Alotaibi, S.; Alqahtani, S.; Alzahrani, A. Recent Advances in Magnetic Resonance Imaging for the Diagnosis of Liver Cancer. *Diagnostics* **2025**, *15*, 2016. [CrossRef] [PubMed]
15. Anichini, M.; Galluzzo, A.; Danti, G.; Grazzini, G.; Pradella, S.; Treballi, F.; Bicci, E. Focal Lesions of the Liver and Radiomics: What Do We Know? *Diagnostics* **2023**, *13*, 2591. [CrossRef] [PubMed]
16. Wang, D.C.; Jang, H.-J.; Kim, T.K. Characterization of indeterminate liver lesions on CT and MRI with contrast-enhanced ultrasound: What is the evidence? *Am. J. Roentgenol.* **2020**, *214*, 1295–1304. [CrossRef] [PubMed]
17. Azizaddini, S.; Mani, N. *Liver Imaging*; StatPearls Publishing: Treasure Island, FL, USA, 2023. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK557460/> (accessed on 15 March 2026).
18. Kahraman, G.; Haberal, K.M.; Dilek, O.N. Imaging features and management of focal liver lesions. *World J. Radiol.* **2024**, *16*, 139. [CrossRef] [PubMed]
19. Noreikaite, J.; Albasha, D.; Chidambaram, V.; Arora, A.; Katti, A. Indeterminate liver lesions on gadoteric acid-enhanced magnetic resonance imaging of the liver: Case-based radiologic-pathologic review. *World J. Hepatol.* **2021**, *13*, 1079–1097. [CrossRef] [PubMed]
20. Mavilia, M.G.; Pakala, T.; Molina, M.; Wu, G.Y. Differentiating cystic liver lesions: A review of imaging modalities, diagnosis and management. *J. Clin. Transl. Hepatol.* **2018**, *6*, 208–216. [CrossRef] [PubMed]
21. Kim, Y.Y.; Lee, S.; Shin, J.; Son, W.J.; Roh, Y.H.; Hwang, J.A.; Lee, J.E. Diagnostic performance of CT versus MRI Liver Imaging Reporting and Data System category 5 for hepatocellular carcinoma: A systematic review and meta-analysis of comparative studies. *Eur. Radiol.* **2022**, *32*, 6723–6729. [CrossRef] [PubMed]
22. Fowler, K.J.; Shetty, A.S.; Brunt, E.M. Radiologic–pathologic correlation of liver tumors. In *Abdominal Radiology Fundamentals*; StatPearls Publishing: Treasure Island, FL, USA, 2021. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK569803/> (accessed on 15 March 2026).
23. Sharma, P.; Yadav, S.; Needham, C.M.; Feuerstadt, P. An updated review of cystic hepatic lesions. *Clin. Res. Hepatol.* **2019**, *5*, 22–29. [CrossRef] [PubMed]
24. Marrero, J.A.; Ahn, J.; Rajender Reddy, K. ACG clinical guideline: The diagnosis and management of focal liver lesions. *Am. J. Gastroenterol.* **2014**, *109*, 1328–1347. [CrossRef] [PubMed]
25. Mortelé, K.J.; Ros, P.R. Cystic focal liver lesions in the adult: Differential CT and MR imaging features. *Radiographics* **2001**, *21*, 895–910. [CrossRef] [PubMed]
26. Grazzini, G.; Cozzi, D.; Flammia, F.; Grassi, R.; Agostini, A.; Belfiore, M.P.; Borgheresi, A.; Mazzei, M.A.; Floridi, C.; Carrafiello, G.; et al. Hepatic tumors: Pitfall in diagnostic imaging. *Acta Biomed.* **2020**, *91*, 9–17. [CrossRef] [PubMed]
27. Peltenburg, J.E.; Hosni, A.; Bahij, R.; Boeke, S.; Braam, P.M.; Hall, W.A.; Intven, M.P.W.; Nicosia, L.; Sonke, J.-J.; Witte, M.; et al. Interobserver variation in tumor delineation of liver metastases using magnetic resonance imaging. *Phys. Imaging Radiat. Oncol.* **2024**, *30*, 100592. [CrossRef] [PubMed]
28. Covert, E.C.; Fitzpatrick, K.; Mikell, J.; Kaza, R.K.; Millet, J.D.; Barkmeier, D.; Gemmete, J.; Christensen, J.; Schipper, M.J.; Dewaraja, Y.K. Intra- and inter-operator variability in MRI-based manual segmentation of hepatocellular carcinoma lesions and its impact on dosimetry. *EJNMMI Phys.* **2022**, *21*, 90. [CrossRef] [PubMed]
29. Karademir, I.; Ward, E.; Peng, Y.; Wise, L.; Buckle, C.; Kunnakkam, R.; Oto, A. Measurements of hepatic metastasis on MR imaging: Assessment of interobserver and intersequence variability. *Acad. Radiol.* **2016**, *23*, 132–143. [CrossRef] [PubMed]
30. Khan, S.; Noor, M.N.; Ashraf, I.; Masud, M.I.; Aman, M. Impact of CT Intensity and Contrast Variability on Deep-Learning-Based Lung-Nodule Detection: A Systematic Review of Preprocessing and Harmonization Strategies (2020–2025). *Diagnostics* **2026**, *16*, 201. [CrossRef] [PubMed]
31. Benfares, A.; Mourabiti, A.y.; Alami, B.; Boukansa, S.; Benomar, I.; El Bouardi, N.; Alaoui Lamrani, M.Y.; El Fatimi, H.; Amara, B.; Serraj, M.; et al. Nomogram Based on the Most Relevant Clinical, CT, and Radiomic Features, and a Machine Learning Model to Predict EGFR Mutation Status in Non-Small Cell Lung Cancer. *J. Respir.* **2025**, *5*, 11. [CrossRef]
32. Xiang, K.; Jiang, B.; Shang, D. The overview of the deep learning integrated into the medical imaging of liver: A review. *Hepatol. Int.* **2021**, *15*, 868–880. [CrossRef] [PubMed]

33. Elhaie, M.; Koozari, A.; Koohi, H.; Alqurain, Q.T. Machine learning and deep learning approaches in MRI for quantifying and staging fatty liver disease: A systematic review. *J. Med. Imaging Radiat. Sci.* **2025**, *56*, 102112. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Benfares, A.; Alami, B.; Boukansa, S.; Qjidaa, M.; Benomar, I.; Serraj, M.; Lakhssassi, A.; Jamil, M.O.; Maaroufi, M.; Qjidaa, H. NSCLC EGFR Mutation Prediction via Random Forest Model: A Clinical–CT–Radiomics Integration Approach. *Adv. Respir. Med.* **2025**, *93*, 39. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Zhou, S.K.; Greenspan, H.; Shen, D. Deep learning for medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Bilic, P.; Christ, P.F.; Vorontsov, E.; Ben-Cohen, A.; Kaissis, G.; Szeskin, A.; Jacobs, C.; Mamani, G.E.H.; Chartrand, G.; Lohöfer, F.; et al. The Liver Tumor Segmentation Benchmark (LiTS). *arXiv* **2019**, arXiv:1901.04056.
37. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **2020**, *295*, 328–338. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Salehi, A.W.; Khan, S.; Gupta, G.; Alzu'bi, S.; Albalawi, F.; Almomani, I.; Alsubai, S.; Alruhaili, M. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, and Future Scope. *Sustainability* **2023**, *15*, 5930. [\[CrossRef\]](#)
39. Mukhlif, A.A.; Al-Khateeb, B.; Mohammed, M.A. Incorporating a Novel Dual Transfer Learning Approach for Medical Images. *Sensors* **2023**, *23*, 570. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Alruwaili, M.; Mohamed, M. An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification. *Diagnostics* **2025**, *15*, 551. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Li, S.; Huang, J. ResGDANet: An Efficient Residual Group Attention Neural Network for Medical Image Classification. *Appl. Sci.* **2025**, *15*, 2693. [\[CrossRef\]](#)
42. Hofmann, A.; Klein, M.; Rueter, D.; Sauer, A. A Deep Residual Neural Network for Image Reconstruction in Biomedical 3D Magnetic Induction Tomography. *Sensors* **2022**, *22*, 7925. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Yahya, A.A.; Liu, K.; Hawbani, A.; Wang, Y.; Hadi, A.N. A Novel Image Classification Method Based on Residual Network, Inception, and Proposed Activation Function. *Sensors* **2023**, *23*, 2976. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Mienye, I.D.; Swart, T.G.; Obaido, G.; Jordan, M.; Ilono, P. Deep Convolutional Neural Networks in Medical Image Analysis: A Review. *Information* **2025**, *16*, 195. [\[CrossRef\]](#)
45. Kallipolitis, A.; Revelos, K.; Maglogiannis, I. Ensembling EfficientNets for the Classification and Interpretation of Histopathology Images. *Algorithms* **2021**, *14*, 278. [\[CrossRef\]](#)
46. Rahman, H.; Khan, A.R.; Sadiq, T.; Farooqi, A.H.; Khan, I.U.; Lim, W.H. A Systematic Literature Review of 3D Deep Learning Techniques in Computed Tomography Reconstruction. *Tomography* **2023**, *9*, 2158–2189. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Altaf, A.R.; Altaf, A.; Rehman, F.U. Modified EfficientNet-B0 Architecture Optimized with Quantum-Behaved Algorithm for Skin Cancer Lesion Assessment. *Diagnostics* **2025**, *15*, 3245. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Zhang, Y.; Wang, J.; Gorriz, J.M.; Wang, S. Deep Learning and Vision Transformer for Medical Image Analysis. *J. Imaging* **2023**, *9*, 147. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Zhang, J.; Li, F.; Zhang, X.; Wang, H.; Hei, X. Automatic Medical Image Segmentation with Vision Transformer. *Appl. Sci.* **2024**, *14*, 2741. [\[CrossRef\]](#)
50. Okpala, C.; Yang, X. Convolutional Neural Network–Vision Transformer Architecture with Gated Control Mechanism and Multi-Scale Fusion for Enhanced Pulmonary Disease Classification. *Diagnostics* **2024**, *14*, 2790. [\[CrossRef\]](#) [\[PubMed\]](#)
51. El Asnaoui, K. Design Ensemble Deep Learning Model for Pneumonia Disease Classification. *Int. J. Environ. Res. Public Health* **2021**, *10*, 55–68. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Rajaraman, S.; Antani, S.K. Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs. *Diagnostics* **2020**, *8*, 27318–27326. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Shah, V.; Keniya, R.; Shridharani, A.; Punjabi, M.; Shah, J.; Mehendale, N. Diagnosis of COVID-19 Using CT Scan Images and Deep Learning Techniques. *Emerg. Radiol.* **2021**, *28*, 497–505. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Liu, J.; Guo, H.; He, Y.; Li, H. Vision Transformer-Based Ensemble Learning for Hyperspectral Image Classification. *Remote Sens.* **2023**, *15*, 5208. [\[CrossRef\]](#)
55. Heaton, J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genet. Program. Evolvable Mach.* **2018**, *19*, 305–307. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.