Received: 10 March 2025 | Revised: 7 August 2025 | Accepted: 31 October 2025 | Published online: 17 November 2025

RESEARCH ARTICLE

Medinformatics
2026, Vol. 3(1) 46–55
DOI: 10.47852/bonviewMEDIN52025652

BON VIEW PUBLISHING

Deep Learning Neural Network with Transfer Learning for Liver Cancer Classification

Nibras Mizouri[1,2,*]

[1]*MIRACL Laboratory, University of Sfax, Tunisia*

[2]*Higher Institute of Computer Science and Multimedia, University of Sfax, Tunisia*

Abstract: Hepatocellular carcinoma (HCC), commonly referred to as "primary" liver cancer, represents the fourth leading cause of cancer-related mortality worldwide. Despite its high fatality rate, several studies have highlighted that early detection and accurate diagnosis can significantly reduce disease progression and improve patient survival outcomes. With the rapid increase in the use of imaging modalities such as computed tomography and magnetic resonance imaging, medical specialists are confronted with the challenge of interpreting vast amounts of complex data. Manual classification of cancer through these imaging techniques is not only labor-intensive and time-consuming but also prone to human error, especially when dealing with large, heterogeneous datasets. In this context, computer-aided diagnosis (CAD) systems have emerged as powerful tools to support radiologists in achieving faster and more reliable decision-making. Thus, the primary goal of this effort is to design an advanced CAD framework for the accurate detection of liver cancer by leveraging the potential of deep learning methodologies. Specifically, transfer learning strategies combined with convolutional neural networks are employed to enhance classification performance. Two state-of-the-art pretrained models, VGG-16 and MobileNet-V1, are utilized within the proposed framework to optimize feature extraction and improve diagnostic accuracy. Experimental results demonstrate that the proposed model achieves an impressive classification accuracy of 96% in distinguishing HCC cases. Data collection (i), data preprocessing (ii), and data analysis (iii) are the three fundamental stages that make up the architecture of the system. Overall, this study provides a reliable and efficient approach for liver cancer detection, offering a substantial contribution to early diagnosis strategies and paving the way for improved clinical decision support in oncology.

Keywords: liver tumor, HCC, deep learning, transfer learning, VGG-16, MobileNet

1. Introduction

The International Agency for Research on Cancer estimates that 9.7 million people died from cancer worldwide in 2022, with an estimated 20 million new cases being diagnosed [1]. Hepatocellular carcinoma (HCC), in particular, is one of them that continues to pose a serious threat to world health. HCC caused more than 866,000 new cases and 758,000 deaths globally, making it the third most common cause of cancer-related fatalities and the sixth most often diagnosed malignancy [2–4]. Since cirrhosis and chronic liver disorders are common causes of HCC, early and precise identification is crucial to enhancing treatment results and survival rates.

In clinical practice, radiologists primarily rely on contrast-enhanced computed tomography (CT) or magnetic resonance imaging (MRI) scans to identify and analyze liver lesions. These modalities provide detailed visualizations of liver structures and potential vascular involvement. However, the interpretation of medical images remains largely manual or semi-automated, which is time-consuming, subjective, and prone to error [5]. Consequently, automated and intelligent diagnostic tools are increasingly being sought to improve accuracy, reduce costs, and enhance early detection efficiency.

For the purpose of identifying and categorizing liver cancers, several computer-aided diagnosis (CAD) systems have been developed over the past decade [6–8]. The recent rise of artificial intelligence, particularly deep learning (DL), has opened new horizons in the field of medical imaging. Deep learning, a subfield of machine learning (ML) based on deep neural networks, has shown significant promise in image classification, segmentation, and disease prediction [9, 10].

Numerous studies have demonstrated the efficacy of DL based approaches across different medical domains. For example, Khan et al. [11] proposed a modified MobileNet model with residual skip blocks, achieving 93.2% accuracy for multiclass brain tumor classification.

Khan et al. [12] developed a hybrid deep dense learning method combining transfer learning (TL) with a Vit-L16 transformer architecture for breast cancer diagnosis. This approach utilizes ResNet50, EfficientNetB1, and a proposed ProDense block as the backbone models for training extensive features, achieving an accuracy of 98.08%.

Khan et al. [13] proposed an effective ensemble method for diabetic retinopathy diagnosis with three pretrained models with

*Corresponding author: Nibras Mizouri, MIRACL Laboratory, University of Sfax and Higher Institute of Computer Science and Multimedia, University of Sfax, Tunisia. Email: nibras.mizouri@isims.com.tn; nibras.elmizouri@gmail.com

the best performance named DenseNet169, MobileNetV1, and Xception for diverse feature extraction.

Regarding liver cancer detection, both traditional ML and deep learning models have been explored by utilizing very accurate picture processing techniques such as image segmentation, detection, analysis, and classification [14–17].

More recently, advanced hybrid models have been developed for HCC prediction. One such model employs a three-component architecture (NCA-GA-SVM) that integrates neighborhood component analysis (NCA) for feature extraction, genetic algorithm (GA) for feature optimization and parameter tuning, and support vector machine (SVM) for binary classification. This two-level feature selection strategy achieved a prediction accuracy of 96.36% when validated on 165 HCC patients [18].

Deep learning-based approaches have further evolved to include complex architectures. Zhou et al. [19] introduced a hierarchical convolutional neural network (CNN) framework for the classification of focal liver lesions from multi-phase CT scans, distinguishing between malignant and benign types across 616 nodules.

Another CNN-based CAD system achieved classification accuracies above 98% across various tumor types using venous-phase CT scans [20].

More recent studies have explored sophisticated architectures. Amin et al. [21] proposed an optimized generative adversarial network (GAN) based framework for liver disease identification, integrating ResNet-50 for feature extraction, YOLO-v3 for tumor detection, and InceptionResNet-v2 for segmentation. Their model achieved 95% accuracy and 0.99 mean average precision (mAP).

Lin et al. [22] achieved 90% accuracy using a VGG-16-based CNN applied to combined two-photon excitation fluorescence (TPEF) and second harmonic generation (SHG) images for HCC detection.

In the study conducted by Chou et al. [23], a neural network-based model was developed for detecting and classifying the severity of nonalcoholic fatty liver disease using B-mode ultrasound images. Multiple pretrained CNNs were evaluated, with ResNet-50 v2 achieving 84% accuracy.

Finally, Chen et al. [24] utilized the SENet model to classify HCC differentiation in histopathological images, outperforming VGG16, ResNet50, and other deep models with an accuracy of 95.27%.

A comparative summary of these recent studies and their main characteristics is provided in Table 1.

This table provides an overview of prior studies on liver cancer detection using deep learning, focusing on imaging types, model architectures, feature extraction methods, and reported results.

One of the major challenges in deploying deep learning models for medical image analysis is the need for large, annotated datasets. In the medical domain, data acquisition is not only costly but also subject to strict privacy regulations and anonymization procedures, which further complicates data collection. Additionally, training deep learning models can be time-consuming and necessitate significant computer resources, such as powerful graphics processing units (GPUs).

Given these limitations, any proposed method must not only focus on accuracy and robustness but also take into account the computational cost and scalability.

In this study, we address the problem of liver cancer (HCC) classification and detection under the constraint of having a limited number of annotated medical images per category. The following section describes the dataset characteristics and our proposed method to overcome these challenges.

2. Materials and Proposed Methods

Figure 1 illustrates the workflow of the proposed CAD system for HCC classification and detection, based on TL. The process begins with the acquisition of liver MRI images, which are first subjected to a data processing stage that includes augmentation and annotation.

Table 1. Comparative summary of recent deep learning approaches for liver cancer (HCC) detection and classification

Study/Author	Image Acquisition	Models/Classifier	Results
Książek et al. [18]	165 CT images	NCA-GA-SVM	Accuracy = 96.36% F1-score = 95.52%
Zhou et al. [19]	616 multi-phase computed tomography (CT) images	Convolutional neural network (CNN)	Accuracy = 73.4%
Li et al. [20]	165 CT scans	CNN-based CAD	Accuracy = 98%
Lin et al.[22]	217 TPEF and SHG images	VGG-16	Accuracy = 90%
Chen et al. [24]	444 histopathological images	SENet, VGG16, ResNet50	Accuracy = 95.27% Sensitivity = 99.41% Precision = 92.02% F1-score = 95.11%
Chou et al. [23]	21,855 B-mode US images	VGG 19, ResNet-50 v2, MobileNet v2, Xception, and Inception v2	Accuracy = 0.84% Sensitivity = 0.83% Specificity = 0.94% F1-score = 0.84%
Amin et al. [21]	CT – 3D-IRCADb-01	ResNet-50, YOLO-v3, InceptionResNet-v2	Accuracy = 95% mAP = 0.99%

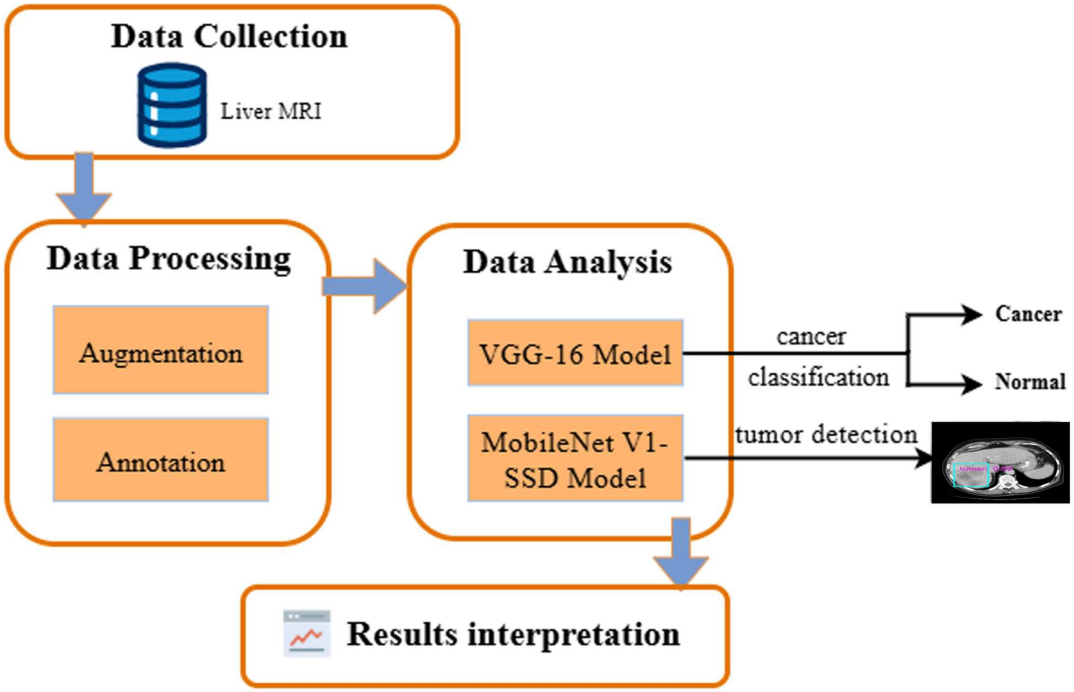

Figure 1. Schema of proposed approach

These processed images are then fed into two deep learning models: VGG-16 for cancer classification and MobileNet-V1-SSD for tumor detection.

The performance of the proposed system is evaluated using standard metrics such as accuracy, sensitivity, and specificity, as discussed in the Results and Discussion section.

2.1. Data collection

The liver MRI images used in this study were provided by the University Hospital of Clermont-Ferrand (CHU), France. All images were acquired by qualified radiologists using standard digital imaging protocols. Prior to image acquisition, informed consent was obtained from each patient, allowing the anonymized data (including medical images and limited clinical metadata) to be used for research purposes. All personal identifiers were removed, and the dataset was securely stored on the CHU's servers in compliance with data protection regulations.

The initial dataset comprised 334 MRI liver images, acquired following the injection of contrast agents to enhance tumor visibility. MRI was selected over CT and ultrasound due to its superior contrast resolution and its ability to provide both morphological and physiological information [25]. The dataset contains both benign (normal) and malignant (cancerous) cases, with image sizes of 520 × 520 and 320 × 260 pixels. 20% of the dataset was put aside for testing, and the remaining 80% was used for training.

2.2. Data processing

2.2.1. Data augmentation

In medical imaging, deep learning models typically perform better when trained on large datasets. However, acquiring large-scale, annotated medical datasets remains challenging due to issues of patient privacy, labeling costs, and the limited availability of expert radiologists. To mitigate these limitations and reduce overfitting, data augmentation techniques are applied to artificially increase the size and diversity of the training dataset [26, 27].

Data augmentation involves applying a range of transformations such as flipping, rotation, noise addition, and blurring while preserving the semantic integrity of the original image. This approach improves the generalization capabilities of the model and helps it adapt better to unseen data.

The specific augmentation parameters used in our study are summarized in Table 2.

Table 2. Data augmentation parameters

Parameter	Description	Value
Flip	Reversing rows or columns of pixels vertically and horizontally	0.3
Rotation	Rotate to the right or left on a 25° axis	25°
Noise	Addition of random noise	0.5
Blur	Application of a light blurring filter	0.1

These parameters were selected based on empirical evaluation and common practices in medical image preprocessing.

We apply a variety of augmentation techniques such as rotation, translation, noise injection, and flip to produce new versions of existing images (Figure 2).

This step played a key role in enhancing the ability of our classification and detection model to generalize by introducing a wide variety of data during its training. Data augmentation was particularly crucial given the limited size of the medical image dataset, making the training data more diverse, which helps to prevent the model from overfitting.

2.2.2. Data annotation

Once the data quantity problem is resolved, an equally important element is needed in the process of our CAD system to classify and detect liver tumors. This element involves assigning semantic labels or textual descriptions to images. In the medical field,

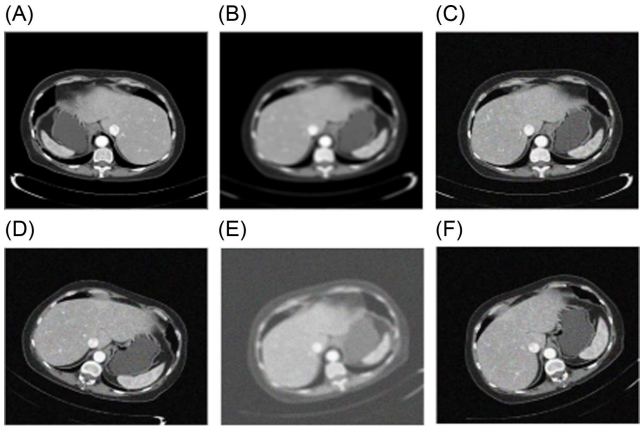

Figure 2. Images augmentation with flip, rotation, noise, and blur after applying dataset generator

assembling an annotated and labeled database presents a challenge because of limited resources.

In addition, it takes a lot of time during the acquisition, which is rarely annotated. Various studies have highlighted this technique [28], which aims to improve medical assistance by creating a help tool that describes, analyzes, and locates such diseases in the content of images.

Effective annotation should ensure two things: first, objects must be accurately labeled with the correct class, and second, pixels that contain the object must be accurately selected.

Annotation in general methods falls into one of three categories:

1) Manual annotation, where the user manually selects and labels the regions of interest.
2) Automatic annotation, where the system automatically detects and labels the semantic content of the images with a set of keywords or using pretrained models.
3) Semi-automatic annotation, which combines both approaches by integrating human assistance into the automatic process.

While manual annotation offers higher precision, it is often labor-intensive and time-consuming. In contrast, automatic methods provide scalability but may suffer from lower accuracy, especially in complex medical cases. Semi-automatic approaches [29, 30] attempt to balance these trade-offs by involving clinicians in a guided annotation process.

In our approach, image annotation is used to detect and localize cancerous lesions in liver MRI scans.

For this task, we employed the LabelImg tool, an open-source graphical annotation tool implemented in Python (Figure 3). It allows for precise manual labeling by selecting the tumor region with bounding boxes.

The annotations are saved in Pascal VOC format (XML files), which are then divided into training and testing subsets for model development and evaluation.

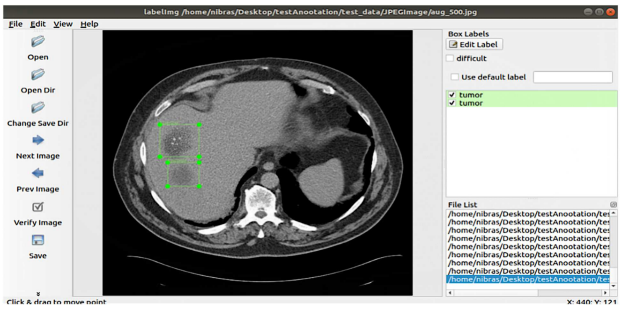

Figure 3. Example image annotation using LabelImg

2.3. Data analysis using transfer learning

The second stage of our proposed CAD system relies on deep learning, employing TL with two well-known models: VGG-16 and MobileNet-V1.

TL [31] is an effective strategy that leverages knowledge gained from solving a source task and applies it to a different but related target task [32, 33]. With the development of deep learning, this approach has been shown to be quite effective.

Indeed, this field's models need a lot of resources and lengthy processing durations. TL efficiently resolves complicated issues with less training time by using pretrained models as a starting point. This enables the quick construction of high-performance models.

Figure 4 shows the contrast between the process of traditional ML and TL. Conventional machine learning, as we can see, attempts to learn each task independently using a distinct model from the beginning. TL, on the other hand, attempts to apply information from earlier source tasks to target activities, which have very little labeled data [34].

Practically, this study aims to implement TL by employing pretrained CNNs for the classification and detection of HCC tumors.

The core idea is straightforward: to reuse and adapt the knowledge embedded in CNN architectures that were originally trained on large-scale datasets (source domain) to effectively address our target task, which involves limited annotated medical images.

In our approach, we utilize two different deep learning architectures: VGG-16, applied for image classification (normal vs. cancerous), and MobileNet-V1-SSD, employed for tumor localization and detection.

2.3.1. VGG-16 model

As shown in Figure 5, the architecture of the VGG-16 [35] model is pretrained with TL.

Originally, this model was trained on the ImageNet dataset, which contains over 14 million natural images categorized into 1,000 classes, and achieved an accuracy of 92.7% [36].

In our proposed classification pipeline, we adopt the pretrained VGG-16 model and modify its final layers to suit the binary classification task of distinguishing between normal and cancerous liver MRI images. Specifically, the original fully connected (FC) layers

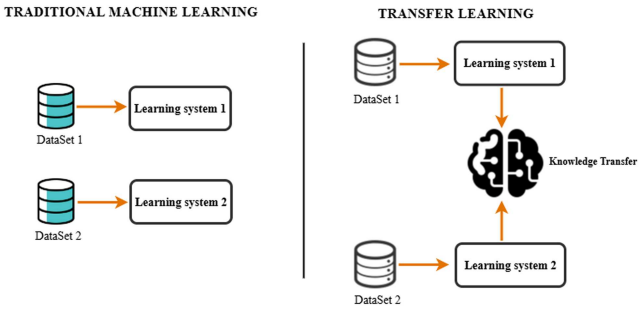

Figure 4. Comparison between transfer learning vs. machine learning

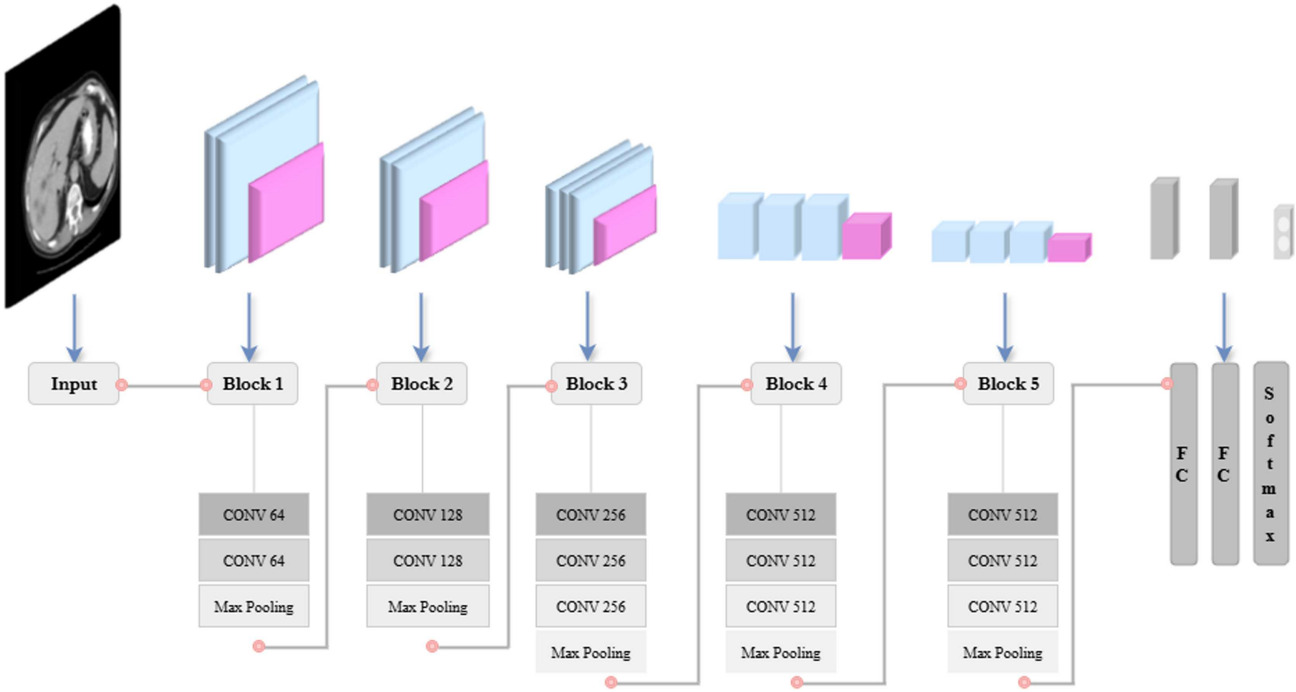

Figure 5. VGG-16 architecture

are replaced with custom layers matching our dataset's number of features and target classes (2 classes).

The VGG-16 architecture consists of 13 convolutional layers and 3 FC layers, structured as follows:

The input to the network is a 224 × 224 × 3 image.

The first 2 convolutional layers use 64 filters, followed by a max pooling layer. Then, the next 2 convolutional layers apply 128 filters, also followed by a max pooling layer.

The subsequent 3 convolutional layers (fifth, sixth, and seventh) use 256 filters, again followed by pooling.

Therefore, the final 6 convolutional layers use 512 filters, grouped in 2 blocks, each ending with max pooling.

Finally, the last classification block includes 2 FC layers with 512 units each and an output layer with 2 neurons, corresponding to the normal and cancer classes.

All hidden layers use the ReLU activation function, while the output layer uses softmax for class probability estimation.

2.3.2. MobileNet-V1-SDD model

MobileNet is a special convolutional neural network architecture introduced by Google in 2017 [37].

The idea behind this model lies in the use of a new type of convolution, "depthwise separable convolution," which drastically reduces the number of parameters and computation compared to standard convolutions (as illustrated in Figure 6, which contrasts it with the standard convolution approach).

Figure 7, which presents the typical structure of a depthwise separable convolution block, shows the separation between depthwise and pointwise convolutions with intermediate ReLU6 activation and batch normalization.

This convolutional operation is composed of two separate layers: depthwise convolution, in which a single filter is applied per input channel, and the pointwise convolution (1 × 1), which combines the outputs of the depthwise step [35].

MobileNet also introduces two hyperparameters: the width multiplier and the resolution multiplier, which provide trade-offs between accuracy and efficiency. These parameters regulate the

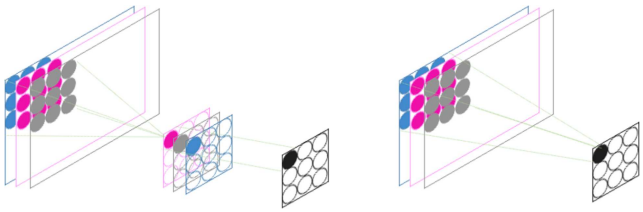

Figure 6. Visualization of depthwise separable structure used in MobileNet (left) and standard convolution (right)

Figure 7. Depthwise separable convolution block

number of input/output channels of convolution layers and the resolution of the input data (i.e., height, width).

For our application, we employ the MobileNet-V1 Single Shot Detector (SSD) framework for the detection and localization of HCC tumors. A one-stage object identification technique called SSD [38] detects objects of varied sizes by using numerous feature maps that have been derived at different scales. In particular, six convolution layers (including conv11 and conv13) are chosen for SSD level detection, and the MobileNet-V1 basis is expanded with eight further convolution layers beyond conv13 (Figure 8).

Each convolutional block in MobileNet is followed by batch normalization and the ReLU6 activation function [39], ensuring the stability and nonlinearity. The final SSD detection layers apply classification and localization predictions across multiple scales,

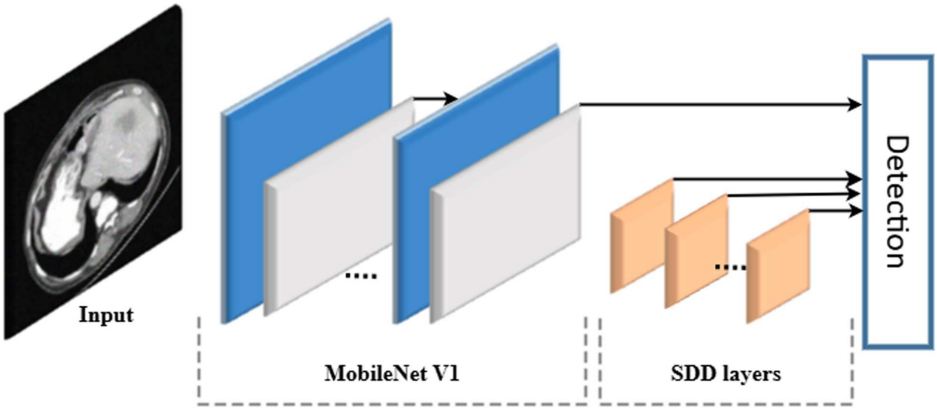

Figure 8. Overall architecture of MobileNet-V1-SSD used for liver tumor detection

enabling the model to identify both small and large tumors in liver MRI images.

The advantages of MobileNet-V1-SSD include its low memory footprint, high detection speed, and suitability for limited resource environments – factors that are especially important in clinical settings.

Table 3 summarizes the main training hyperparameters used for VGG-16 and MobileNet-V1-SSD models, including learning rates, batch sizes, optimizers, and number of epochs.

Table 3. Parameter values for models

Hyperparameter	VGG-16	MobileNet-V1-SSD
Input image size	224 × 224	300 × 300
Learning rate	0.001	0.001 (base), 0.01 (head)
Epochs	100	50
Batch size	8	5
Optimizer	Adam	SGD

3. Results and Discussion

We conducted all our tests and training on a laptop equipped with an Intel Core i7-7500U CPU/3.50 GHz processor with 8 GB RAM and an NVIDIA GeForce GT920M graphics card.

The CAD method presented consists of two steps: the preparation of data and the training of models used for the classification and detection of cancerous lesions. During data preparation, augmentation and manual annotation were employed to enhance dataset quality and diversity. The training phase involved two deep learning models: VGG-16 for classification (normal vs. cancerous) and MobileNet-V1-SSD for tumor detection and localization.

A total of 2000 liver MRI images were used through data augmentation techniques (see Section 2.2), with 80% allocated for training and 20% for testing. This augmentation was critical to improving model generalization and reducing overfitting.

3.1. Classification performance with VGG-16

The proposed model was trained for 100 epochs. As shown in Figure 9, from epoch 0 to epoch 29, the training accuracy increases rapidly, reaching approximately 92%, and then gradually converges to around 96%. Similarly, a significant drop in the training loss is observed from epoch 0 to epoch 25, decreasing to approximately 0.12. After that, the loss stabilizes around 0.09. This behavior indicates that the model learns efficiently during the initial epochs and then reaches a stable performance, suggesting convergence with minimal overfitting.

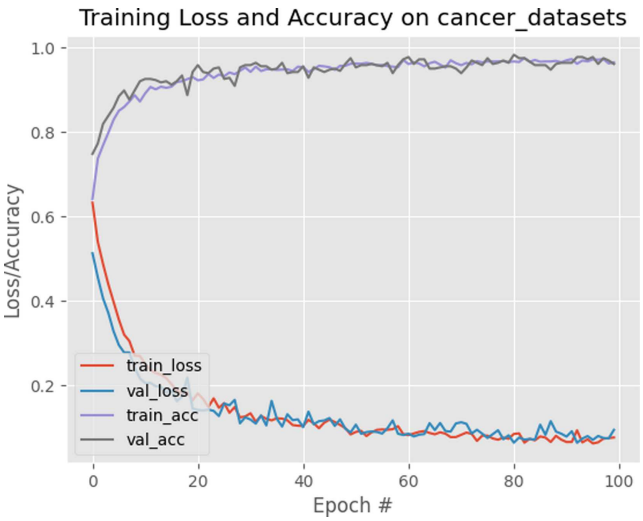

Figure 9. Training and validation accuracy/loss of the proposed model VGG-16

In addition, Table 4 provides a detailed evaluation of the model's predictive capability on the test set. Specifically, 173 of the 184 HCC cases were correctly classified as cancer (true positives), while 11 were misclassified as normal (false negatives). Similarly, out of 182 normal cases, 179 were accurately identified as normal (true negatives), whereas only 3 were incorrectly labeled as cancer (false positives). These results support the model's robustness and generalization ability when applied to new data.

True positive refers to correctly identified tumor pixels, true negative refers to correctly identified non-tumor pixels, false

Table 4. Confusion matrix for the classification model

		Predicted values	
		Cancer	Normal
Actual values	Cancer	173	11
	Normal	3	179

positive refers to incorrectly identified tumor pixels, and false negative refers to incorrectly identified non-tumor pixels.

Each of these metrics is computed for the two classes, and the average of all these metrics is used to determine the algorithm's overall measure.

Accuracy, confusion matrix, precision, and F1-score are the criteria used to analyze the performance of the suggested model, and they are displayed in Table 5.

Table 5. Evaluation metrics

Classifier	VGG-16
Accuracy (%)	96
Sensitivity (%)	94
Specificity (%)	98
Precision (%)	98
F1-score (%)	98

The performance measures considered are defined as follows:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \tag{3}$$

$$\text{F1-score} = 2 \times \frac{\text{Prcision} \times \text{Sensitivity}}{\text{Prcision} + \text{Sensitivity}} \tag{4}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

where TP, FP, TN, and FN indicate the number of true positives, false positives, true negatives, and false negatives, respectively.

The VGG-16 model demonstrated excellent performance on the liver MRI dataset. It achieved an accuracy of 96%, a sensitivity of 94%, and a specificity of 98%, indicating a high ability to correctly classify both cancerous and normal cases (Table 5). Moreover, the precision and F1-score, both at 98%, reflect a balanced performance between detecting true cancer cases and minimizing false positives – an essential criterion in medical image analysis.

3.2. Detection performance with MobileNet-V1-SSD

The detection stage was conducted using the MobileNet-V1-SSD architecture, trained to localize liver tumors within MRI images. Prior to training, a manual annotation process was carried out using the LabelImg tool, where each tumor was labeled with bounding boxes and tagged as "tumor".

The dataset (340 images annotated in.jpg and.xml) is used: 80% for training and 20% for testing.

The model was trained over 50 epochs, with performance evaluated at each step using the PASCAL VOC metric.

The detector's localization and classification performance are assessed using the PASCAL VOC assessment [40]. According to the PASCAL VOC metric (Figure 10), a detection is deemed valid if it has a 50% Jaccard overlap and the same prediction label as the ground truth.

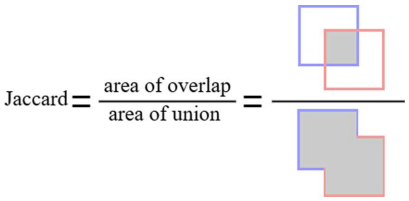

Figure 10. Illustration of Jaccard calculation

As the values FP, FN, and TP can be modified by setting different thresholds, it is crucial to evaluate the accuracy and sensitivity at various thresholds in order to gauge the overall performance of a detector model. mAP is the average of the APs for all classes, whereas average accuracy (AP) is the average of the maximum accuracy at different recalls for each category.

Table 6 presents a snapshot of the training console output during the final epochs (36 to 49) of the MobileNet-V1-SSD model. The table displays the progression of the total validation loss, as well as its components (regression loss and classification loss) across epochs. These logs confirm the consistent trend observed in the quantitative analysis, with a gradual reduction in the loss values indicating improved model performance. Notably, the total validation loss decreased to 2.899 by epoch 49, with a regression loss of 0.972 and classification loss of 1.957.

Table 6. Training log output of MobileNet-V1-SSD

Epoch	Validation loss	Validation regression loss	Validation classification loss
36	3.2228	1.1355	2.8056
37	2.9965	0.9990	2.8056
38	2.9961	0.9990	2.8349
39	3.0714	1.0289	2.1266
40	1.8877	0.0612	2.1266
41	2.5906	1.1291	2.0662
42	2.9436	0.9478	1.9958
43	2.9542	0.9478	1.9958
44	3.0776	1.1097	2.2579
45	3.0655	0.9897	2.8668
46	2.9629	0.9629	1.9598
47	2.8288	0.9496	1.9733
48	2.9292	0.9496	1.9733
49	2.9326	0.9724	1.9597

This table summarizes the results of the MobileNet-V1-SSD model from epoch 36 to 49, showing validation loss, regression loss, and classification loss evolution.

After evaluation on the validation set, the trained model achieved a mAP of 63%, indicating a moderate yet encouraging ability to correctly detect and localize liver tumors within the limited dataset.

These values indicate an acceptable trade-off between localization accuracy and classification confidence. While the mAP is lower than typical classification accuracies, this is expected in

detection tasks, especially in medical contexts (liver tumor localization), particularly with limited and manually annotated data, which significantly impacted model performance. Compared to traditional detectors such as YOLOv3 and Faster R-CNN [41, 42], which often report mAP scores above 70% in large-scale datasets, the MobileNet-SSD model offers a computationally efficient alternative suitable for clinical environments with constrained resources.

Figure 11 shows liver tumor localization using MobileNet-V1-SSD with an overlap score. This metric represents the probability value of the tumor covering the bounding box (for our model, the threshold is set at 50%). In the event that the score is below the cut-off, TN is recognized, and the tumor is not found. For various MRI images, the Figure 11 shows the 74%, 97%, and 92% confidence scores for tumor detection during testing.

3.3. Discussion and limitations

The proposed CAD system, combining VGG-16 for classification and MobileNet-V1-SSD for detection, has demonstrated promising results in liver cancer diagnosis. The classification model achieved an accuracy (96%), reflecting its robustness in differentiating between normal and cancerous liver MRI images. Additionally, the detection model yielded a mAP of 63%, which is considered acceptable given the complexity of medical object detection and the limited size of the dataset.

One of the key strengths of the proposed approach lies in the balance it achieves between computational efficiency and diagnostic performance. The use of TL with pretrained models for HCC prediction significantly reduced the execution time and led to improved accuracy.

However, several limitations must be acknowledged:

One major limitation is the requirement for large and high-quality datasets. Deep learning including TL heavily relies on vast amounts of well-curated data to learn patterns and make accurate classifications and localizations. Although data augmentation increased the dataset to 2000 images, the number of annotated HCC datasets MRI (approximately 334) remains relatively small for training models due to the rarity of the disease and the need for comprehensive clinical and imaging data. This scarcity may limit the model's ability to generalize to unseen data or diverse patient populations.

Annotation quality and consistency are crucial in tumor detection, and this is another limitation. Manual annotation using LabelImg introduced the potential for human error and inconsistency, particularly for small or low-contrast tumors. The quality of bounding boxes directly influences the detection performance of the MobileNet-V1-SSD.

While the classification model achieved excellent results, the detection model's mAP of 63% suggests room for improvement (moderate detection accuracy). This score indicates that some tumors, particularly those with atypical appearance or boundary locations, were missed or incorrectly localized.

In addition, the generalizability of deep learning models and the imbalanced representation of tumor types are other limitations. The dataset may not fully represent all variations in liver tumor types (e.g., stages, shapes, intensity), which could bias the model toward more common cases and reduce its effectiveness in rare or complex scenarios, leading to inaccurate predictions.

4. Conclusion and Future Work

This paper presents a deep learning-based CAD system for liver tumor classification and detection in MRI images. In this system two powerful pretrained models: VGG-16 for classification and MobileNet-V1-SSD for tumor localization which a set of image augmentation and annotation techniques employed to prepare the data.

The system demonstrated high classification performance (96% accuracy) and acceptable detection results (63% mAP), highlighting its potential for clinical support in liver cancer diagnosis.

The results indicate that TL, coupled with data augmentation and manual annotation, can effectively mitigate the challenges posed by small annotated medical datasets. Future research should focus on improving accuracy in a number of areas, including integrating volumetric (3D) data to better understand tumors spatially and investigating the combination of clinical data with other modalities, such as genetic information, imaging data (MRI, CT scans), video data, and blood-based biomarkers. Additionally, adding geographically varied data to the dataset will guarantee generalizability and prevent overfitting, improve the assessment of changing risk profiles, and make it possible to identify high-risk individuals early.

Account for the presence of other chronic conditions as diabetes or hepatitis that may influence HCC development. Develop

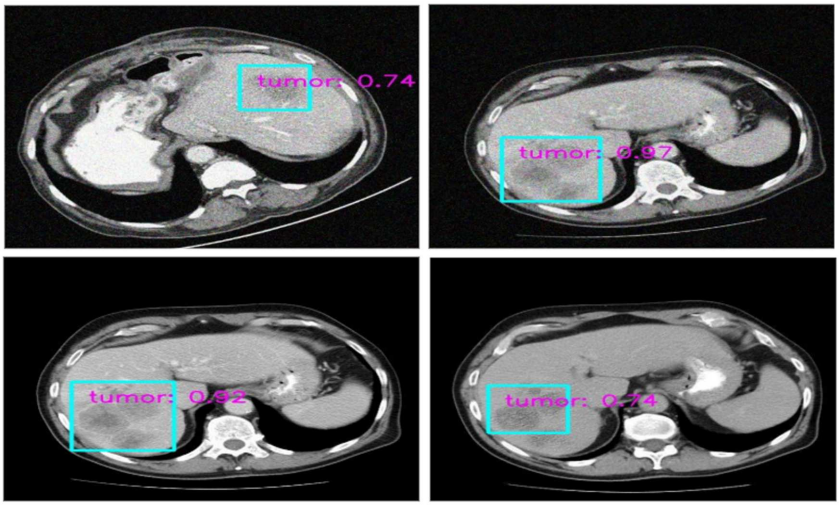

Figure 11. Example liver tumor localization results using MobileNet-V1-SSD

models that aim to integrate data capturing changes in clinical variables over time in order to better assess evolving risk profiles and enable the early identification of high-risk individuals.

In short, focusing on these future work directions with incorporating semi-automatic annotation methods may also improve localization performance.

Acknowledgment

The authors acknowledge that an earlier version of this manuscript was published as a preprint at "Research Square" (https://doi.org/doi.org/10.21203/rs.3.rs-2355564/v1) on December 12, 2022. The preprint had not undergone peer review. This published article incorporates revisions made during the peer review process for *Medinformatics*.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

The data that support this work are available upon reasonable request to the corresponding author.

Author Contribution Statement

Nibras Mizouri: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. https://doi.org/10.3322/caac.21834

[2] Chon, Y. E., Jeong, S. W., & Jun, D. W. (2021). Hepatocellular carcinoma statistics in South Korea. *Clinical and Molecular Hepatology*, *27*(3), 512–514. https://doi.org/10.3350/cmh.2021.0171

[3] McGlynn, K. A., Petrick, J. L., & El-Serag, H. B. (2021). Epidemiology of hepatocellular carcinoma. *Hepatology*, *73*, 4–13. https://doi.org/10.1002/hep.31288

[4] Oh, J. H., & Jun, D. W. (2023). The latest global burden of liver cancer: A past and present threat. *Clinical and Molecular Hepatology*, *29*(2), 355–357. https://doi.org/10.3350/cmh.2023.0070

[5] Yang, J. D., & Heimbach, J. K. (2020). New advances in the diagnosis and management of hepatocellular carcinoma. *BMJ*, *371*, m3544. https://doi.org/10.1136/bmj.m3544

[6] Kalejahi, B. K., Meshgini, S., Danishvar, S., & Khorram, S. (2022). Diagnosis of liver disease by computer-assisted imaging techniques: A literature review. *Intelligent Data Analysis*, *26*(4), 1097–1114. https://doi.org/10.3233/IDA-216379

[7] Liu, X., Cruz Rivera, S., Moher, D., Calvert, M. J., Denniston, A. K., & SPIRIT-AI and CONSORT-AI Working Group. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *The Lancet Digital Health*, *2*(10), e537–e548. https://doi.org/10.1016/S2589-7500(20)30218-1

[8] Trivedi, N. K., Tiwari, R. G., Anand, A., Gautam, V., Witarsyah, D., & Misra, A. (2022). Application of machine learning for diagnosis of liver cancer. In *2022 International Conference Advancement in Data Science, E-learning and Information Systems*, 1–5. https://doi.org/10.1109/ICADEIS56544.2022.10037379

[9] Afza, F., Sharif, M., Khan, M. A., Tariq, U., Yong, H.-S., & Cha, J. (2022). Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. *Sensors*, *22*(3), 799. https://doi.org/10.3390/s22030799

[10] Hekmat, A., Zuping, Z., Bilal, O., & Khan, S. U. R. (2025). Differential evolution-driven optimized ensemble network for brain tumor detection. *International Journal of Machine Learning and Cybernetics*, *16*(9), 6447–6472. https://doi.org/10.1007/s13042-025-02629-6

[11] Khan, S. U. R., Asim, M. N., Vollmer, S., & Dengel, A. (2025). *AI-driven diabetic retinopathy diagnosis enhancement through image processing and salp swarm algorithm-optimized ensemble network.* arXiv. https://doi.org/10.48550/ARXIV.2503.14209

[12] Khan, S. U. R., Asif, S., & Bilal, O. (2025). Ensemble architecture of vision transformer and CNNs for breast cancer tumor detection from mammograms. *International Journal of Imaging Systems and Technology*, *35*(3), e70090. https://doi.org/10.1002/ima.70090

[13] Khan, S. U. R., Zhao, M., & Li, Y. (2025). Detection of MRI brain tumor using residual skip block based modified MobileNet model. *Cluster Computing*, *28*(4), 248. https://doi.org/10.1007/s10586-024-04940-3

[14] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: A friendly introduction. *Journal of Big Data*, *9*(1), 102. https://doi.org/10.1186/s40537-022-00652-w

[15] Khan, R. A., Fu, M., Burbridge, B., Luo, Y., & Wu, F.-X. (2023). A multi-modal deep neural network for multi-class liver cancer diagnosis. *Neural Networks*, *165*, 553–561. https://doi.org/10.1016/j.neunet.2023.06.013

[16] Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S., & Alzahrani, A. (2022). A deep learning approach for liver and tumor segmentation in CT images using ResUNet. *Bioengineering*, *9*(8), 368. https://doi.org/10.3390/bioengineering9080368

[17] Zhang, Y., Lv, X., Qiu, J., Zhang, B., Zhang, L., Fang, J., ..., & Zhang, S. (2021). Deep learning with 3D convolutional neural network for noninvasive prediction of microvascular invasion in hepatocellular carcinoma. *Journal of Magnetic Resonance Imaging*, *54*(1), 134–143. https://doi.org/10.1002/jmri.27538

[18] Książek, W., Turza, F., & Pławiak, P. (2022). NCA-GA-SVM: A new two-level feature selection method based on neighborhood component analysis and genetic algorithm in hepatocellular carcinoma fatality prognosis. *International Journal for Numerical Methods in Biomedical Engineering*, *38*(6), e3599. https://doi.org/10.1002/cnm.3599

[19] Zhou, J., Wang, W., Lei, B., Ge, W., Huang, Y., Zhang, L., ..., & Wang, W. (2021). Automatic detection and classification of focal liver lesions based on deep convolutional neural networks: A preliminary study. *Frontiers in Oncology*, *10*, 581210. https://doi.org/10.3389/fonc.2020.581210

[20] Li, J., Wu, Y., Shen, N., Zhang, J., Chen, E., Sun, J., ..., & Zhang, Y. (2020). A fully automatic computer-aided diagnosis

system for hepatocellular carcinoma using convolutional neural networks. *Biocybernetics and Biomedical Engineering*, *40*(1), 238–248. https://doi.org/10.1016/j.bbe.2019.05.008

[21] Amin, J., Anjum, M. A., Sharif, M., Kadry, S., Nadeem, A., & Ahmad, S. F. (2022). Liver tumor localization based on YOLOv3 and 3D-semantic segmentation using deep neural networks. *Diagnostics*, *12*(4), 823. https://doi.org/10.3390/diagnostics12040823

[22] Lin, H., Wei, C., Wang, G., Chen, H., Lin, L., Ni, M., ..., & Zhuo, S. (2019). Automated classification of hepatocellular carcinoma differentiation using multiphoton microscopy and deep learning. *Journal of Biophotonics*, *12*(7), e201800435. https://doi.org/10.1002/jbio.201800435

[23] Chou, T.-H., Yeh, H.-J., Chang, C.-C., Tang, J.-H., Kao, W.-Y., Su, I.-C., ..., & Su, E. C.-Y. (2021). Deep learning for abdominal ultrasound: A computer-aided diagnostic system for the severity of fatty liver. *Journal of the Chinese Medical Association*, *84*(9), 842–850. https://doi.org/10.1097/JCMA.0000000000000585

[24] Chen, C., Chen, C., Ma, M., Ma, X., Lv, X., Dong, X., . . . , & Chen, J. (2022). Classification of multi-differentiated liver cancer pathological images based on deep learning attention mechanism. *BMC Medical Informatics and Decision Making*, *22*(1), 176. https://doi.org/10.1186/s12911-022-01919-1

[25] Gravina, M., Marrone, S., Sansone, M., & Sansone, C. (2021). DAE-CNN: Exploiting and disentangling contrast agent effects for breast lesions classification in DCE-MRI. *Pattern Recognition Letters*, *145*, 67–73. https://doi.org/10.1016/j.patrec.2021.01.023

[26] Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., & Shen, F. (2022). *Image data augmentation for deep learning: A survey*. arXiv. https://doi.org/10.48550/ARXIV.2204.08610

[27] Zhou, W., Wang, H., & Wan, Z. (2022). Ore image classification based on improved CNN. *Computers and Electrical Engineering*, *99*, 107819. https://doi.org/10.1016/j.compeleceng.2022.107819

[28] Hameed, I. M., Abdulhussain, S. H., & Mahmmod, B. M. (2021). Content-based image retrieval: A review of recent trends. *Cogent Engineering*, *8*(1), 1927469. https://doi.org/10.1080/23311916.2021.1927469

[29] Aljabri, M., AlAmir, M., AlGhamdi, M., Abdel-Mottaleb, M., & Collado-Mesa, F. (2022). Towards a better understanding of annotation tools for medical imaging: A survey. *Multimedia Tools and Applications*, *81*(18), 25877–25911. https://doi.org/10.1007/s11042-022-12100-1

[30] Haider, T., & Michahelles, F. (2021). Human-machine collaboration on data annotation of images by semi-automatic labeling. In *Proceedings of Mensch und Computer 2021*, 552–556. https://doi.org/10.1145/3473856.3473993

[31] Pan, S. J. (2014). Transfer learning. In C. C. Aggarwal (Ed.), *Data classification: Algorithms and applications* (1st ed., pp. 536-570). Chapman and Hall/CRC.

[32] Iman, M., Arabnia, H. R., & Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies*, *11*(2), 40. https://doi.org/10.3390/technologies11020040

[33] Kapoor, A., Gulli, A., & Pal, S. (2022). *Deep learning with TensorFlow and Keras: Build and deploy supervised, unsupervised, deep, and reinforcement learning models* (3rd ed.). UK: Packt Publishing.

[34] Ayana, G., Dese, K., & Choe, S.-W. (2021). Transfer learning in breast cancer diagnoses via ultrasound imaging. *Cancers*, *13*(4), 738. https://doi.org/10.3390/cancers13040738

[35] Yang, H., Ni, J., Gao, J., Han, Z., & Luan, T. (2021). A novel method for peanut variety identification and classification by improved VGG16. *Scientific Reports*, *11*(1), 15756. https://doi.org/10.1038/s41598-021-95240-y

[36] Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. *BMC Medical Imaging*, *22*(1), 69. https://doi.org/10.1186/s12880-022-00793-7

[37] Yu, G., Wang, L., Hou, M., Liang, Y., & He, T. (2020). An adaptive dead fish detection approach using SSD-MobileNet. In *2020 Chinese Automation Congress* , 1973–1979. https://doi.org/10.1109/CAC51589.2020.9326648

[38] Kumar, A., Zhang, Z. J., & Lyu, H. (2020). Object detection in real time based on improved single shot multi-box detector algorithm. *EURASIP Journal on Wireless Communications and Networking*, *2020*(1), 204. https://doi.org/10.1186/s13638-020-01826-x

[39] Liu, Z., Xiang, X., Qin, J., Tan, Y., Zhang, Q., & Xiong, N. N. (2021). Image recognition of citrus diseases based on deep learning. *Computers, Materials & Continua*, *66*(1), 457–466. https://doi.org/10.32604/cmc.2020.012165

[40] Li, W. (2021). Analysis of object detection performance based on Faster R-CNN. *Journal of Physics: Conference Series,1827* (1), 012085. https://doi.org/10.1088/1742-6596/1827/1/012085

[41] Islam, S. U., Ferraioli, G., Pascazio, V., Vitale, S., & Amin, M. (2024). Performance analysis of YOLOv3, YOLOv4 and MobileNet SSD for real time object detection. *The Sciencetech*, *5*(2), 37–49.

[42] Yamagishi, S., Doman, K., Mekada, Y., Nishida, N., & Kudo, M. (2022). Detection and tracking of liver tumors for ultrasound diagnostic support using deep learning. *Journal of Image and Graphics*, *10*(1), 50–55. https://doi.org/10.18178/joig.10.1.50-55

How to Cite: Mizouri, N. (2026). Deep Learning Neural Network with Transfer Learning for Liver Cancer Classification. *Medinformatics*, *3*(1), 46–55. https://doi.org/10.47852/bonviewMEDIN52025652