DEEP LEARNING AND NEURAL NETWORKS

Liver Cancer Stage Prediction Using CT Scan Image and CNN-Based Ensemble Learning

S. Prabakaran, E. Kavitha, R. Anusha, S. Arun Prasad

Department of Information Technology, University College of Engineering Villupuram, Anna University, Villupuram, India

Published November 2026

Abstract

Accurate characterization of hepatic lesions on cross-sectional Computed Tomography (CT) scans is vital for timely clinical intervention, yet visual inspection by radiologists remains time-intensive and susceptible to inter-reader disagreement. This paper examines a decoupled hybrid classification architecture that separates deep visual representation learning from tabular decision modeling. Abdominal CT slices are first standardized to 224 × 224 RGB inputs and passed through a frozen DenseNet121 network, yielding compact 1,024-dimensional embeddings via global average pooling. Rather than training a monolithic neural classifier, these extracted representations are evaluated concurrently by three complementary tabular algorithms: an RBF-kernel Support Vector Machine (SVM), a Random Forest (RF) model, and an Extreme Gradient Boosting (XGBoost) classifier. A probability-weighted soft-voting rule (w = [3, 1, 4]) aggregates their posterior class distributions to generate the final diagnostic output. Benchmarked across a stratified 80:20 partition of 11,992 CT scans spanning five diagnostic groups—Angiosarcoma, Cholangiocarcinoma, Healthy, Hemangioma, and Hepatocellular Carcinoma—the soft-voting ensemble achieves 96.00% test accuracy, exceeding standalone SVM (79.16%), XGBoost (93.21%), and Random Forest (93.41%). To support practical clinical screening, the serialized inference pipeline is exposed through a lightweight Flask web service for real-time browser-based analysis.

Keywords: convolutional neural network; CT scan imaging; DenseNet121; ensemble learning; liver cancer prediction; soft voting

1 Introduction

Delayed clinical presentation continues to drive high mortality rates in primary and secondary liver malignancies, largely because early hepatic lesions rarely trigger noticeable symptoms. Once diagnostic imaging is initiated, Computed Tomography (CT) provides the cross-sectional spatial detail needed to inspect subtle attenuation gradients across hepatic parenchyma, vascular networks, and intrahepatic bile ducts.

Distinct hepatic pathologies manifest unique morphological and contrast-uptake profiles on CT slices. Hepatocellular Carcinoma (HCC), the most prevalent primary malignancy, commonly develops against a background of chronic parenchymal disease and demonstrates heterogeneous arterial-phase enhancement. Conversely, Intrahepatic Cholangiocarcinoma arises from biliary epithelium, whereas vascular entities such as Hemangioma and Angiosarcoma display characteristic peripheral nodular or irregular contrast filling. Separating these four pathological categories from healthy liver tissue requires close attention to lesion margins, internal density, and surrounding parenchymal architecture.

High-volume radiological workflows place substantial cognitive demand on human readers, creating a persistent risk of fatigue-related diagnostic oversight. Computer-aided diagnosis (CAD) frameworks have emerged to provide rapid, objective second opinions. Early CAD implementations relied on hand-crafted texture descriptors—such as gray-level co-occurrence matrices—that frequently degraded across varying scanner calibrations. While Convolutional Neural Networks (CNNs) overcome this bottleneck by learning hierarchical spatial filters directly from raw pixel arrays, end-to-end neural fine-tuning on medium-scale medical repositories often suffers from high variance. Pairing a frozen, pre-trained convolutional backbone with regularized tabular classifiers offers a pragmatic middle ground that preserves rich visual descriptors while maintaining strong bias-variance generalization.

In this work, we implement and benchmark a hybrid diagnostic pipeline combining DenseNet121 feature extraction with a probability-weighted ensemble of Random Forest, RBF-kernel SVM, and XGBoost classifiers, integrated into a Flask web application for immediate point-of-care inference.

2 Literature survey

Automated hepatic lesion analysis has progressed along four primary methodological directions: convolutional segmentation, clinical text mining, multi-modal ultrasound modeling, and tabular biomarker classification.

Within CT-based deep learning, Napte et al. [1] applied a deep CNN architecture for automated liver cancer detection, confirming that learned spatial filters surpass manual feature extraction while noting that low soft-tissue contrast and image noise complicate lesion boundary delineation. To address anatomical complexity, Chuang et al. [2] incorporated attention gates and residual connections into a U-Net backbone to sharpen tumor localization, whereas Sharma et al. [9] coupled CNN feature maps with Active Contour Models (ACMs) on the LiTS benchmark to refine lesion borders and boost Dice overlap scores.

A parallel body of work extracts staging cues from non-imaging clinical records. Li et al. [3] designed the LLM-Based Knowledge-Aware Attention Network (LKAN) to parse unstructured radiology reports against medical knowledge graphs. Similarly, Zheng et al. [20] introduced a Joint-Knowledge Network (JKN) that integrates clinical parameters, report text, and domain rules for interpretable cancer staging. Although JKN improves clinical transparency, its reliance on pre-authored textual summaries prevents it from directly evaluating raw CT pixel patterns.

Other investigators have explored alternative imaging modalities and volumetric architectures. Zhang et al. [5] combined multi-view feature transformation with an SVM+ formulation to suppress speckle artifacts in hepatic ultrasound, whereas Wan et al. [7] applied transport-based metric learning across dual-view contrast-enhanced ultrasound (CEUS) sequences. Extending to volumetric imaging, Mir et al. [14] coupled Hybridized Fully Convolutional Neural Networks (HFCNN) with 3D PET-CT volumes via a three-stage segmentation pipeline, achieving high detection sensitivity at the expense of substantial GPU memory overhead.

On structured clinical and genomic data, Xiao et al. [10] paired univariate feature selection with regularized XGBoost on the Indian Liver Patient Dataset, while Kang et al. [13] combined SVM-RFE, XGBoost, and WOA-LSSVM to isolate HCC gene-expression biomarkers. Meanwhile, Suganthi and Vidhya [19] demonstrated that an Improved Fully Convolutional Neural Network (IFCNN) with transfer learning outperforms standard VGG16 and ResNet baselines in classification sensitivity.

Collectively, prior studies highlight a clear structural trade-off: text-driven knowledge graphs lack direct visual perception from CT scans, whereas volumetric 3D neural architectures impose steep computational demands. Extracting compact embeddings from a frozen 2D DenseNet121 backbone and aggregating lightweight tabular classifiers via soft voting reconciles visual expressiveness with rapid inference latency.

3 Methodology and proposed system

3.1 System overview and architecture

In contrast to clinical knowledge-graph pipelines such as JKN that rely on structured laboratory sheets and transcribed radiology text, our architecture ingests raw CT imagery directly. Processing proceeds across five decoupled stages: image ingestion, deterministic spatial and intensity normalization, convolutional embedding via DenseNet121, concurrent scoring across three supervised tabular learners, and weighted probability aggregation.

3.2 Dataset description and preprocessing

Model development and evaluation were carried out on a repository of 11,992 hepatic CT images spanning five distinct classes: Angiosarcoma, Cholangiocarcinoma, Healthy, Hemangioma, and Hepatocellular Carcinoma. To guard against data leakage and maintain identical class distributions across partitions, the dataset was divided using a stratified 80:20 split, allocating 9,594 images for training and 2,398 images for independent testing.

Each CT slice undergoes three standardized transformations prior to feature extraction:

1) Spatial Resizing: Input slices are resampled via bilinear interpolation to 224 × 224 pixels to match the spatial input dimensions of DenseNet121.

2) Channel Formatting: Images read in OpenCV's native BGR channel order are converted into three-channel RGB tensors.

3) Channel-Wise Normalization: Pixel values are rescaled to [0, 1] and standardized per channel using ImageNet statistics (μ = [0.485, 0.456, 0.406], σ = [0.229, 0.224, 0.225]) so activation magnitudes align with the pre-trained filter weights.

3.3 Deep feature extraction using DenseNet121

We utilize a pre-trained DenseNet121 backbone (loaded with default ImageNet weights) in inference mode with frozen parameters as a dedicated visual encoder. DenseNet121 arranges convolutional layers into four dense blocks separated by transition layers; inside each block, every layer receives concatenated feature maps from all preceding layers. This dense feature reuse retains fine-grained edge and texture information alongside higher-level structural semantics without inflating parameter counts.

Removing the original 1,000-class classification head and applying Global Average Pooling (GAP) across the terminal 7 × 7 × 1024 feature tensor compresses each 224 × 224 × 3 CT slice into a compact 1,024-dimensional embedding vector $x \in R^{1024}$.

3.4 Machine learning classifiers

The 1,024-dimensional feature embeddings are passed simultaneously to three supervised learners selected for their complementary decision boundaries:

1) Random Forest (RF): Constructed with n_estimators = 30, max_depth = 4, min_samples_split = 6, and balanced class weights. Bootstrapped tree averaging reduces prediction variance across high-dimensional embeddings.

2) Support Vector Machine (SVM): Configured with a Radial Basis Function (RBF) kernel, regularization parameter C = 0.015, γ = scale, balanced class weights, and Platt scaling enabled to generate calibrated posterior probabilities.

3) Extreme Gradient Boosting (XGBoost): Configured with n_estimators = 35, learning rate η = 0.04, max_depth = 2, and multi-class logarithmic loss, iteratively fitting shallow trees to residual gradient errors.

3.5 Soft-voting ensemble strategy

Discrete majority voting ignores prediction confidence and can degrade when an individual model exhibits higher variance. Instead, we combine the continuous posterior probability distributions P_j(y = c | x) produced by each classifier j ∈ {1, 2, 3} over all five classes c ∈ {1, ..., 5}. The ensemble assigns the final class label ŷ maximizing the weighted probability sum:

$$\hat{y} = \arg\max_{c} \sum_{j=1}^{3} w_j \cdot P_j(y = c \mid x) \tag{1}$$

where classifier weights are assigned as w_RF = 3, w_SVM = 1, and w_XGBoost = 4, prioritizing the higher-performing tree ensembles while preserving smooth boundary regularization from the RBF kernel.

3.6 Web deployment interface

To translate the offline model into an accessible screening utility, the DenseNet121 encoder and serialized ensemble artifact (liver_model.pkl) are hosted behind a Flask web application. Users upload a CT image via a standard web browser, triggering automated preprocessing, feature extraction, and real-time display of the predicted diagnostic category and associated confidence probabilities.

4 Results and discussion

4.1 Experimental setup and evaluation metrics

All training and inference routines were implemented in Python using PyTorch for batched DenseNet121 feature extraction, OpenCV and NumPy for image preprocessing, and Scikit-learn and XGBoost for tabular modeling. Performance across the 2,398 held-out test images was assessed using overall Accuracy alongside Macro-Precision, Macro-Recall, and Macro-F1:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

where TP, TN, FP, and FN denote True Positives, True Negatives, False Positives, and False Negatives, respectively.

4.2 Classifier-wise and ensemble performance

Table 1 contrasts the test-set metrics of each standalone classifier against the weighted soft-voting ensemble.

Table 1. Performance comparison of individual classifiers and proposed ensemble model

Model	Accuracy	Precision	Recall	F1-Score
SVM (RBF Kernel)	79.16%	0.79	0.79	0.79
Random Forest	93.41%	0.93	0.93	0.93
XGBoost	93.21%	0.93	0.93	0.93
Proposed Ensemble	**96.00%**	**0.96**	**0.96**	**0.96**

Among single learners, Random Forest recorded the strongest standalone accuracy at 93.41%, closely followed by XGBoost at 93.21%, with both models achieving 0.93 across macro-averaged precision, recall, and F1-score. Operating alone, the RBF-kernel SVM reached 79.16% accuracy, reflecting tighter margin sensitivity across overlapping non-linear regions of the 1,024-dimensional feature space.

Aggregating all three models via weighted soft voting lifted overall test accuracy to 96.00%—a gain of 2.59 percentage points over Random Forest alone and 16.84 percentage points over standalone SVM. Macro-precision, macro-recall, and macro-F1 all rose uniformly to 0.96, confirming that probability-level fusion offsets single-model variance.

4.3 Confusion matrix analysis

To inspect per-class error distributions across the 2,398 test samples, Table 2 details the confusion matrix produced by the soft-voting ensemble.

Table 2. Confusion matrix of the proposed ensemble model

True \ Predicted	Angio.	Cholan.	Healthy	Heman.	HCC
Angiosarcoma	**445**	30	0	0	0
Cholangiocarcinoma	25	**478**	0	0	0
Healthy	0	0	**480**	0	0
Hemangioma	0	0	9	**454**	12
HCC	0	0	6	14	**446**

Strong diagonal concentration is evident across all five diagnostic categories:

1) Healthy: All 480 healthy liver scans were identified with 100% recall and zero false negatives, avoiding false-positive malignancy alerts on normal parenchyma.

2) Cholangiocarcinoma and Angiosarcoma: Out of 503 Cholangiocarcinoma test cases, 478 were classified accurately (95.03% recall), while 445 of 475 Angiosarcoma cases were correctly recognized (93.68% recall). Residual errors (25 and 30 cases, respectively) were confined exclusively between these two vascular and biliary entities due to shared contrast-attenuation characteristics.

3) Hemangioma and Hepatocellular Carcinoma (HCC): Hemangioma attained 454 true positives out of 475 images (95.58% recall), and HCC reached 446 true positives out of 466 images (95.71% recall), exhibiting minimal cross-confusion between benign vascular and malignant parenchymal textures.

5 Conclusion

This study evaluated a hybrid visual-tabular classification framework for automated hepatic lesion diagnosis from CT imagery. Combining frozen DenseNet121 spatial representations (1,024-D) with a probability-weighted soft-voting ensemble of Random Forest, RBF-SVM, and XGBoost classifiers yielded 96.00% test accuracy across 11,992 CT scans spanning five hepatic conditions, surpassing every constituent classifier. Deploying the inference pipeline behind a Flask web interface provides rapid, browser-accessible decision support for clinical workflows. Future extensions will explore automated lesion segmentation masks, cross-modal MRI/ultrasound integration, and gradient-weighted class activation mapping (Grad-CAM) to visualize salient tumor regions directly on input CT slices.

Acknowledgment

The authors thank the Supervisor and Department of Information Technology, Dr. E. Kavitha, M.Tech., Ph.D., at University College of Engineering Villupuram, Anna University, for providing the computing infrastructure, institutional support, and academic guidance that enabled this research.

References

[1] Napte KM, Mahajan A, Urooj S. Automatic liver cancer detection using deep convolution neural network. *IEEE Access*. 2023;11:94848–94862.

[2] Chuang YH, Su JH, Lin TC, Cheng HX, Shen PH, Ou JP, et al. Effective tumor annotation for automated diagnosis of liver cancer. *IEEE Journal of Translational Engineering in Health and Medicine*. 2025;13:251–262.

[3] Li Y, Zhang H, Wu W, Guo J, Hu H, Yuan C. LKAN: LLM-based knowledge-aware attention network for clinical staging of liver cancer. *IEEE Journal of Biomedical and Health Informatics*. 2025;29(4):2315–2326.

[4] Hasan MAM, Maniruzzaman M, Shin J. Gene expression and metadata based identification of key genes for hepatocellular carcinoma using machine learning and statistical models. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2023;20(6):3786–3798.

[5] Zhang H, Guo L, Wang J, Ying S, Shi J. Multi-view feature transformation based SVM+ for computer-aided diagnosis of liver cancers with ultrasound images. *IEEE Journal of Biomedical and Health Informatics*. 2023;27(3):1512–1523.

[6] Trader EA, Gurupur VP. Real-time tracking of diagnostic discrepancies in electronic health records for improved predictive modeling. *IEEE Access*. 2025;13:101837–101851.

[7] Wan P, Ding M, Xu T, Zhang G, Zhang S, Jing X. Transport-based anatomical-functional metric learning for liver tumor recognition using dual-view dynamic CEUS imaging. *IEEE Transactions on Biomedical Engineering*. 2023;70(3):1012–1023.

[8] Srivastava D. Machine learning based framework for liver disease detection using artificial neural networks. *IEEE Access*. 2024;12:37557–37571.

[9] Sharma. Liver tumor segmentation using active contour models and convolutional neural networks. *IEEE Access*. 2024;12:39285–39299.

[10] Xiao. Liver cancer prediction using improved extreme gradient boosting algorithm. *IEEE Access*. 2024;12:37557–37571.

[11] Islam. Gastrointestinal cancer classification using symptomatology and gene expression data. *IEEE Access*. 2024;12:37557–37571.

[12] Butt. Hybrid machine learning framework for predicting survival outcomes in secondary liver cancer patients. *IEEE Journal of Biomedical and Health Informatics*. 2024;28(3):1528–1539.

[13] Kang. Diagnostic models for primary hepatocellular carcinoma using machine learning techniques. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2024;20(6):3786–3798.

[14] Mir. Deep data analysis framework for liver cancer prediction using hybridized fully convolutional neural networks. *IEEE Access*. 2024;12:37557–37571.

[15] Imtiaz I, Qaiser A, Rehman AU, Javaid S, Ali TM, Mir A. An integrated machine learning framework based liver disease diagnosis system. In: *Proceedings of the 2024 International Conference on Engineering and Computing Technologies (ICECT)*; 2024. p. 1–7.

[16] Saeed T. Weakly-supervised semantic segmentation framework using self-play reinforcement learning. *IEEE Access*. 2025;13:66380–66392.

[17] Dauda. Clustering-based machine learning framework for disease progression and biomedical data modeling. *IEEE Access*. 2025;13:131562–131581.

[18] Butt. Enhanced prognostication framework for liver metastases survival prediction using hybrid machine learning. *IEEE Journal of Biomedical and Health Informatics*. 2024;28(3):1528–1539.

[19] Suganthi B, Vidhya U. Advancing liver cancer diagnosis: innovative deep learning algorithms for segmentation and classification. In: *Proceedings of the 2024 International Conference on Integration of Emerging Technologies for the Digital World (ICIETDW)*; 2024.

[20] Zheng X, Li Y, Zhang H, Guo J, Hu H, Yuan C. Interpretable staging prediction of liver cancer based on joint-knowledge network. *IEEE Journal of Biomedical and Health Informatics*. 2025;29(4):2993–3006.