electronics

MDPI

Article

nnShuffleNet: Combining a Degrouped 3D ShuffleNet Backbone with a PBDL Head for Hepatocellular Carcinoma Classification from 3D CT

Jiaqi Zhao and Yuexiang Li *

Faculty of Data Science, City University of Macau, Macau SAR, China
* Correspondence: yuexiang.li@ieee.org

Abstract

Three-dimensional (3D) CT classification of hepatocellular carcinoma (HCC) is challenging under conditions of limited cohort size and constrained computational resources. In this study, we construct a framework termed nnShuffleNet and empirically evaluate it. This framework couples a degrouped 3D ShuffleNet backbone with a fixed 50-dimensional principal component analysis and an ordinary least squares classification head (PCA50-OLS). The study cohort comprises 268 non-duplicated CT volumes from 54 independent patients, with 27 HCC and 27 non-HCC patients. The primary evaluation adopts patient-exclusive five-fold cross-validation: all volumes belonging to the same patient are consistently assigned to the same data partition in every fold, and patient-level scores are obtained by averaging the predicted probabilities over all volumes of that patient. nnShuffleNet achieved a five-fold mean patient-level AUC of 0.769 ± 0.125 (patient-bootstrap 95% confidence interval: 0.635–0.883); the same degrouped backbone with a fully connected (FC) head yielded 0.719 ± 0.078, and Swin Transformer + FC yielded 0.763 ± 0.147. In the volume-level comparison on the identical held-out outer test folds of unseen patients, nnShuffleNet attained an AUC of 0.748 ± 0.115. Compared with our implementation of 3D ResNet-50, nnShuffleNet reduced the parameter count and the counted Conv3d/Linear multiply–accumulate operations by 28.46% and 45.37%, respectively. These results support the combination of a degrouped 3D ShuffleNet with PCA50-OLS as an empirical baseline that balances discriminative performance and computational cost under the present cohort and evaluation protocol.

Keywords: hepatocellular carcinoma; 3D CT; nnShuffleNet; patient-exclusive validation; deep learning; medical image classification

Check for updates

Academic Editor: Ferdinando Di Martino

Received: 1 September 2026
Revised: 17 November 2026
Accepted: 23 November 2026
Published: 26 November 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

Liver cancer is the sixth most common cancer and the fourth leading cause of cancer death globally. Without effective interventions, new cases and deaths are projected to double by 2050 [1]. Hepatocellular carcinoma (HCC) accounts for approximately 80% of all primary liver cancers [1–3]. Recent multicenter studies have explored the role of artificial intelligence in the detection, diagnosis, and image classification of focal liver lesions [4,5], providing a foundation for automated analysis using CT information. For the 3D classification of HCC and non-HCC cases, models need to utilize spatial information within the volume while adapting to limited data scale and practical computational conditions. Therefore, classification performance and model complexity should be jointly examined under a consistent and explicit evaluation setting.

In recent years, deep learning networks have achieved remarkable progress in medical image analysis [6]. In resource-constrained clinical environments, lightweight architectures and efficient training strategies have attracted increasing attention [7]. In HCC-related research, Sarfati et al. combined 3D CT deep features with handcrafted radiological features to assist classification [8]; Chang et al. proposed ReViT, utilizing 3D CT for BCLC staging within a multiple instance learning framework [9]; Parimala et al. explored the combination of radiomics features and 3D-CNN in liver lesion classification [10]. Firat and Hanbay's research on hyperspectral images further provided a cross-task example of 3D convolution for joint representation [11]. Additionally, liver lesion classification in multi-phase CT has introduced designs such as dual-resolution Transformers and phase-related information fusion [12,13].

HCC classification from 3D CT images is clinically meaningful but challenging, as available datasets are typically small and computational budgets in practical deployment may be limited. Recent 3D deep learning methods tend to prioritize larger backbones and increasingly complex classification heads. However, for small medical datasets, increased architectural complexity does not necessarily lead to better robustness or deployment efficiency. Thus, under small-sample conditions, can increasing the complexity of the backbone or classification head stably translate into benefits on independent patients?

Based on ShuffleNet and its three-dimensional (3D) extension [14,15], we construct a network framework termed nnShuffleNet. The proposed method comprises two complementary components. First, we remove the grouping constraint in the spatial convolutions of the 3D ShuffleBlock to increase cross-channel connections. Subsequently, after the backbone is trained, we introduce a classification head termed the PCA-Bias-Driven Linear (PBDL) head, which fits a fixed 50-dimensional principal component analysis (PCA), an explicit intercept, and an ordinary least squares (OLS) classification mapping using the features from the training folds.

The main contributions of this work are summarized as follows:

1. We curate an institutional cohort comprising 268 non-duplicate 3D CT volumes from 54 patients, given the challenges of acquiring 3D medical data and the scarcity of such datasets, we consider the scale of this dataset an important contribution.
2. We propose an efficient and effective deep learning framework, no-new-ShuffleNet (nnShuffleNet), specifically designed for HCC classification. This architecture enhances cross-channel information flow by removing grouped convolutions, thereby improving the original 3D ShuffleNet backbone, and introduces a PCA-Bias-Driven Linear (PBDL) classification head.

2. Materials and Methods

2.1. Overall Framework

Our nnShuffleNet consists of a degrouped 3D ShuffleNet backbone and a PCA50-OLS classification head. A single-channel 3D CT volume is processed by the backbone followed by global average pooling, yielding a 960-dimensional feature vector. The backbone is first jointly trained with a fully connected (FC) classification head. After selecting a checkpoint, its parameters are frozen, and the PCA and OLS parameters are fitted using the features from the training folds.

During inference, the input volume is passed through the fixed backbone to obtain features, which are then mean-centered using the training-set mean, projected by the PCA fitted on the training set, and mapped through a linear layer with an intercept to obtain two-class logits. Classification probabilities are finally obtained via the softmax function. The framework is evaluated as an integrated combination: the backbone is responsible for 3D representation learning, while the closed-form linear head is used to examine whether

a compact classification mapping on frozen features can complement a conventional FC head under small-cohort conditions.

This framework is designed to address two practical questions. First, can a degrouped 3D backbone maintain strong classification performance for hepatocellular carcinoma (HCC)? Second, after freezing this backbone, can a classification head composed of a fixed 50-dimensional PCA, an explicit intercept, and ordinary least squares (OLS) fitted solely on training-fold features—namely, the PBDL head—achieve volume-level macro-average AUC scores that are comparable to, or even better than, those of a conventional fully connected (FC) head, thereby demonstrating that a simple, interpretable PCA-based linear classification head can serve as an adequate alternative to a complex FC head?

2.2. Degrouped ShuffleBlock

The 3D ShuffleNet implementation used in this study is based on the ShuffleNet branch structure and its 3D extension [14,15]. Both the standard and degrouped versions share the same stage widths and 960-dimensional output. The main difference lies in the spatial $3 \times 3 \times 3$ convolutions: the standard implementation uses per-channel grouping, while the degrouped version sets the `groups` of the corresponding convolutions to 1, allowing spatial convolutions to utilize information from different channels simultaneously. The remaining stage organization and output dimensions are kept consistent.

The motivation for removing grouping constraints is to strengthen cross-channel interactions of volumetric features. As shown in Figure 1a, to enrich the extracted features, the proposed degrouped ShuffleBlock evenly splits the input feature x into x_1 and x_2 and feeds them into two distinct branches. In each branch, all grouping constraints on the $3 \times 3 \times 3$ convolutions are removed to promote information exchange between channels. Meanwhile, the $1 \times 1 \times 1$ convolutions and channel shuffle operations are preserved to maintain computational efficiency. This modification aims to strengthen channel fusion and improve feature separability without resorting to heavier backbone families.

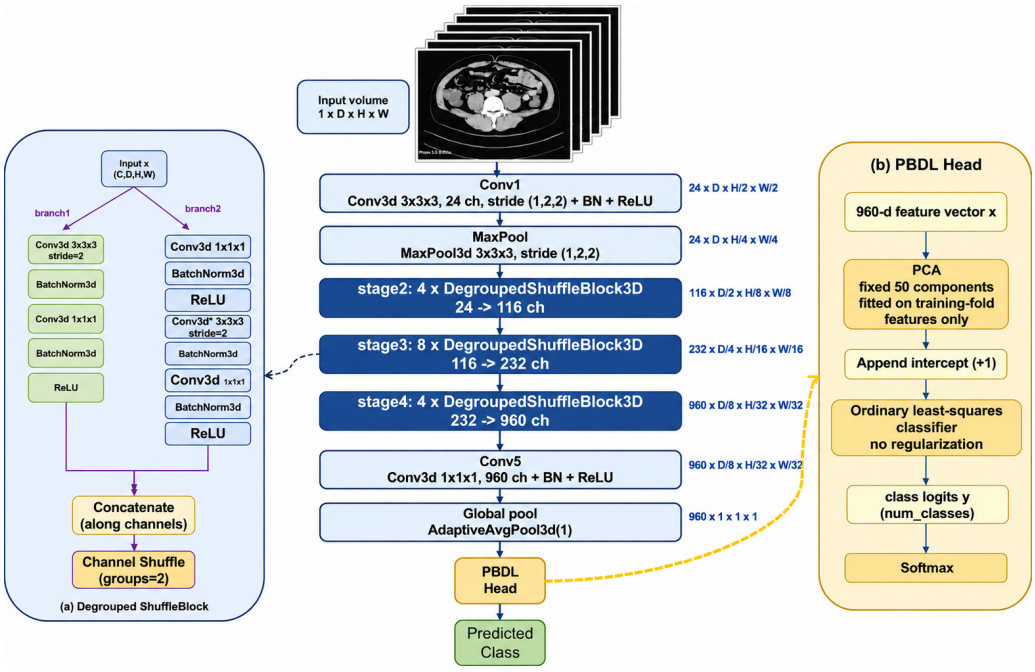


Figure 1. The nnShuffleNet architecture. (**a**) The degrouped 3D ShuffleBlock; * indicates the degrouped convolution, i.e., group partitioning is removed and standard convolution is used, while the number of channels remains unchanged; (**b**) the main PCA50-OLS head, comprising a fixed 50-component PCA projection, an explicit intercept, and ordinary least-squares fitting without regularization.

2.3. PCA-Bias-Driven Linear Head (PBDL)

To construct a classifier on frozen backbone features without additional end-to-end backpropagation, we adopt a PBDL head combining PCA dimensionality reduction, explicit bias, and ordinary least squares. Let the training feature matrix X contain n samples and $d = 960$ features, with a training feature mean of μ. PCA is fitted only on training features without whitening; the main setting requests 50 principal components, and the actual dimension is $k = \min(50, n, d)$. Centering and projection are formulated as:

$$Z = (X - 1\mu^T)V_k, \quad k = \min(50, n, d) \tag{1}$$

where V_k is the PCA loading matrix. To explicitly fit the bias, a column of ones is appended to the reduced features to form the design matrix A; the label matrix Y uses one-hot encoding for HCC and non-HCC classes. Classification parameters are obtained via least squares:

$$A = [Z|1], \quad \Theta^* = \arg\min_{\Theta} \|A\Theta - Y\|_F^2 = A^{\dagger}Y \tag{2}$$

The last row of Θ^* is taken as the bias b, and the remaining rows as W. Test features first undergo the training-set PCA transformation, then compute the two-class logits and softmax scores:

$$\ell(x) = (x - \mu^T)V_k W + b, \quad p_c(x) = \frac{\exp(\ell_c(x))}{\sum_{j=0}^{1} \exp(\ell_j(x))} \tag{3}$$

The bias is applied only once during prediction; original label 0 corresponds to HCC, and the patient-level HCC evaluation uses the 0-th column score. PBDL fitting is performed after the backbone is frozen, so no further updates to backbone parameters are needed for this head. For determined PCA loadings and OLS coefficients, centering, projection, and linear mapping can be combined into a single affine layer.

We choose PCA50-OLS primarily for three reasons. First, it reduces the 960-dimensional backbone representation to the same 50-dimensional space in every fold. Second, the augmented OLS design contains 51 columns (50 components plus the intercept), fewer than the 151–191 training volumes available in the five folds, so the fitted system remains overdetermined. Third, OLS introduces no ridge-penalty hyperparameter that would require an additional search in this small cohort. PCA is fitted exclusively on training-fold features; neither validation nor test features contribute to the PCA basis or OLS coefficients.

2.4. Training Strategy and FC-PBDL Collaboration

Training is divided into two stages: backbone optimization and head fitting. First, the backbone and FC head are jointly trained, selecting checkpoints based on validation volume macro F1; then, the backbone is frozen, and fixed features are extracted from training, validation, and test volumes. PBDL's PCA and linear parameters are fitted only on training features, and test data are used only for final evaluation.

nnShuffleNet and all trainable baselines were implemented in PyTorch version 2.8, which is the release immediately preceding the latest version. and trained on an NVIDIA RTX A6000 GPU (NVIDIA Corporation, Santa Clara, CA, USA). Each backbone was initialized randomly and optimized for 50 epochs using AdamW with a learning rate of 10^{-4}, weight decay of 10^{-4}, batch size 4, and focal loss with $\gamma = 2$ [16]. No learning-rate scheduler was used. For each outer fold, the checkpoint with the highest validation volume-level macro F1 was retained. The test partition was not used for checkpoint selection or hyperparameter fitting.

Input volumes were resized to 64 × 64 × 64 using first-order index-grid interpolation and then normalized by the per-volume z-score. Training augmentation comprised random flipping along two planar directions and random Gaussian noise; validation, testing, and frozen-feature extraction used no random augmentation. This resizing did not constitute isotropic resampling based on physical voxel spacing.

2.5. Dataset

This study uses an institutional dataset of HCC and non-HCC cases. The initial archive contained 401 CT-volume entries from 54 patients. During the data audit, we identified 133 fully duplicate file entries associated with 27 patients. These entries were redundant archived copies rather than independent CT acquisition data. All duplicate files were removed prior to data splitting; no unique CT sequences or patients were excluded. The final cohort therefore contained **268 non-duplicate 3D CT volumes from 54 independent patients**: 27 HCC patients with 135 volumes and 27 non-HCC patients with 133 volumes. Each volume corresponds to a 3D CT reconstruction series in the source archive; multiple series from the same patient are treated as correlated observations. The non-HCC category included pathologically confirmed hepatitis, cholangiocarcinoma, and other benign liver lesions and should not be interpreted as a healthy-control group. Figure 2 shows representative image samples from our dataset.

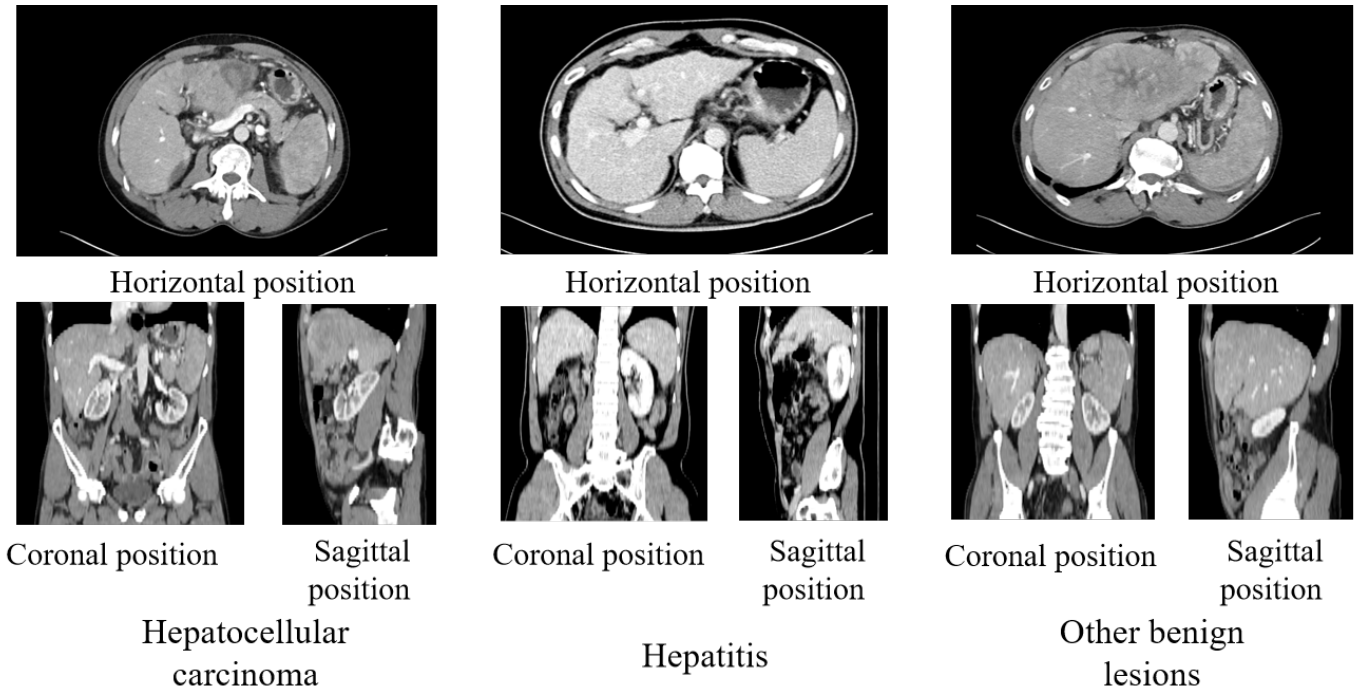


Figure 2. Example samples of hepatocellular carcinoma and non-hepatocellular carcinoma (including hepatitis and other benign lesions) from our HCC dataset.

All reported predictive results use patient-exclusive five-fold cross-validation. Within every outer fold, the training, validation, and test patient identities are mutually exclusive, and all volumes belonging to one patient remain in the same partition. Each of the 54 patients appears in an outer test set exactly once. The exact anonymized fold assignments are provided in Supplementary Table S1.

2.6. Patient-Exclusive Five-Fold Validation

The primary evaluation unit is the patient. For outer test patient i, the HCC probability is the arithmetic mean of the HCC probabilities from that patient's n_i volumes:

$$p_i^{\text{patient}} = \frac{1}{n_i} \sum_{j=1}^{n_i} p_{ij}^{\text{volume}}. \quad (4)$$

Patient-level AUC and macro F1 are calculated separately in each outer test fold and reported as the arithmetic mean ± population standard deviation across the five folds [17,18]. The secondary volume-level analysis uses individual volumes from the same outer test folds; thus, its test volumes also originate exclusively from patients unseen during training and validation. The term "volume-level" denotes the evaluation unit and does not imply a random split by volume.

The five outer-fold compositions are listed in Table 1 ; anonymized patient assignments, class labels, and per-patient volume counts are supplied in Supplementary Table S1.

Table 1. Composition of the five patient-exclusive outer folds. The HCC/non-HCC count refers to test patients.

Fold	Train Patients/Volumes	Validation Patients/Volumes	Test Patients/Volumes	Test Patients HCC/Non-HCC
1	34/176	9/37	11/55	5/6
2	34/184	9/38	11/46	5/6
3	34/151	9/51	11/66	6/5
4	34/162	9/47	11/59	3/8
5	35/191	9/35	10/42	8/2

2.7. Evaluation Metrics and Statistical Criteria

The sample-level benchmark reports the area under the receiver operating characteristic curve (AUC), sensitivity (Sens), and specificity (Spec) [17,18]. Patient-level and volume-level AUCs are computed as the macro-average of two "one-vs-rest" class-specific AUCs; because the two softmax probabilities are complements in this binary classification task, taking the complement jointly for class labels and scores yields identical numerical values for the HCC and non-HCC class-specific AUCs. Consequently, this macro-average simplifies to a single class-specific AUC.

We also employed 5000 stratified patient bootstrap replicates to estimate the uncertainty of the patient-level mean AUC. Within each outer fold and within each class, patients were resampled with replacement, after which the five fold-specific AUCs were recalculated and averaged. Paired model comparisons used the same bootstrap samples. Two-sided bootstrap p-values were obtained by doubling the smaller proportion of pairwise AUC differences that were less than zero or greater than zero, and the Holm–Bonferroni correction procedure was applied to the seven predefined comparisons against nnShuffleNet. This analysis directly targeted the reported five-fold average patient AUCs, rather than pooled uncalibrated probabilities across folds from independent training runs. It is worth noting that all comparison models shared the same patient fold splits, input preprocessing, training duration, checkpoint selection rules, and patient-level aggregation pipeline.

2.8. Baseline Methods

The baseline backbones were standard 3D ShuffleNet [14,15], 3D ResNet-50 [19], 3D EfficientNet [20], 3D DenseNet-121 [21], 3D ViT [22], and 3D Swin Transformer [23]. FC classifiers were evaluated for every backbone. PCA50-OLS was additionally evaluated

on the degrouped ShuffleNet, ResNet-50, and DenseNet-121 representations. All models reported in Tables 2–4 use the same patient folds, input preprocessing, training duration, and checkpoint-selection rule. For configurations appearing in both Tables 2 and 4, the patient-level and volume-level results differ only in the evaluation unit.

Table 2. Patient-level results under the common patient-exclusive five-fold protocol. Values are fold mean ± population SD. Bold indicates the highest mean in each metric; all metrics are higher-is-better. ↑ indicates higher-is-better.

Configuration	Patient AUC ↑	Macro F1
nnShuffleNet (Ours)	**0.769 ± 0.125**	0.585 ± 0.138
Swin Transformer + FC	0.763 ± 0.147	0.600 ± 0.121
ResNet-50 + PBDL	0.755 ± 0.100	0.587 ± 0.149
DenseNet-121 + PBDL	0.744 ± 0.103	0.605 ± 0.146
Degrouped ShuffleNet + FC	0.719 ± 0.078	**0.615 ± 0.116**
DenseNet-121 + FC	0.712 ± 0.113	0.592 ± 0.138
Standard ShuffleNet + FC	0.709 ± 0.120	0.433 ± 0.045
ViT + FC	0.703 ± 0.178	0.599 ± 0.274
ResNet-50 + FC	0.673 ± 0.089	0.568 ± 0.122
EfficientNet + FC	0.608 ± 0.145	0.509 ± 0.095

Table 3. Fold-stratified patient-bootstrap uncertainty and paired comparisons with nnShuffleNet. Differences are nnShuffleNet minus the listed comparator; 5000 paired bootstrap replicates were used. Holm adjustment covers the seven comparisons in this table.

Comparator	Comparator AUC	95% CI	ΔAUC	Raw p	Holm p
nnShuffleNet (ours)	0.769	0.635–0.883	—	—	—
Degrouped ShuffleNet + FC	0.719	0.567–0.852	+0.050	0.356	1.000
Standard ShuffleNet + FC	0.709	0.552–0.848	+0.060	0.447	1.000
ResNet-50 + PBDL	0.755	0.615–0.880	+0.014	0.827	1.000
DenseNet-121 + PBDL	0.744	0.588–0.880	+0.025	0.753	1.000
Swin Transformer + FC	0.763	0.620–0.888	+0.006	0.933	1.000
ViT + FC	0.703	0.533–0.860	+0.066	0.446	1.000
EfficientNet + FC	0.608	0.437–0.767	+0.161	0.045	0.314

Table 4. Volume-level results on the same unseen-patient outer test folds. Values are fold mean ± population SD. Sens. is HCC sensitivity; Spec. is non-HCC specificity. Bold indicates the highest mean in each metric; all metrics are higher-is-better. ↑ indicates higher-is-better.

Configuration	AUC ↑	Sens. ↑	Spec. ↑
nnShuffleNet (Ours)	**0.748 ± 0.115**	0.644 ± 0.232	0.676 ± 0.269
Degrouped ShuffleNet + FC	0.701 ± 0.060	0.596 ± 0.225	**0.720 ± 0.248**
ResNet-50 + FC	0.694 ± 0.098	**0.714 ± 0.171**	0.487 ± 0.078
Standard ShuffleNet + FC	0.723 ± 0.109	0.593 ± 0.281	0.575 ± 0.272
DenseNet-121 + FC	0.671 ± 0.109	0.705 ± 0.245	0.527 ± 0.080
ViT + FC	0.680 ± 0.098	0.584 ± 0.213	0.650 ± 0.342
EfficientNet + FC	0.634 ± 0.107	0.611 ± 0.289	0.546 ± 0.269
Swin Transformer + FC	0.712 ± 0.088	0.693 ± 0.190	0.500 ± 0.148

3. Results

3.1. Patient-Exclusive Five-Fold Performance

We present the primary patient-level results in Table 2. We observed that our proposed nnShuffleNet achieved the highest numerical mean AUC, 0.769 ± 0.125, followed by Swin Transformer+FC (0.763 ± 0.147), ResNet-50+PCA50-OLS (0.755 ± 0.100), and

DenseNet-121+PCA50-OLS (0.744 ± 0.103). On the same degrouped backbone, we found that PCA50-OLS increased the mean patient AUC of our nnShuffleNet from 0.719 to 0.769. For patient-level macro F1, Degrouped ShuffleNet + FC achieved the highest mean value of 0.615, whereas nnShuffleNet achieved 0.585.

We report the patient-bootstrap interval and paired comparison results explicitly in Table 3. Our proposed nnShuffleNet had a mean patient AUC with a 95% CI of 0.635–0.883. We include exact raw and Holm-adjusted p-values to make the uncertainty analysis reproducible.

We visualize the five-fold patient-level ROC and precision–recall (PR) curves for seven representative configurations in Figure 3. The ROC panel shows nnShuffleNet with the highest five-fold mean AUC among the displayed models (0.769), followed by Swin Transformer with FC (0.763). The PR panel reports average precision (AP) as a separate measure; its displayed AP is highest for Swin Transformer with FC (0.867), while nnShuffleNet has an AP of 0.835. These curves summarize fold-specific test predictions from patients absent from the corresponding training and validation partitions.

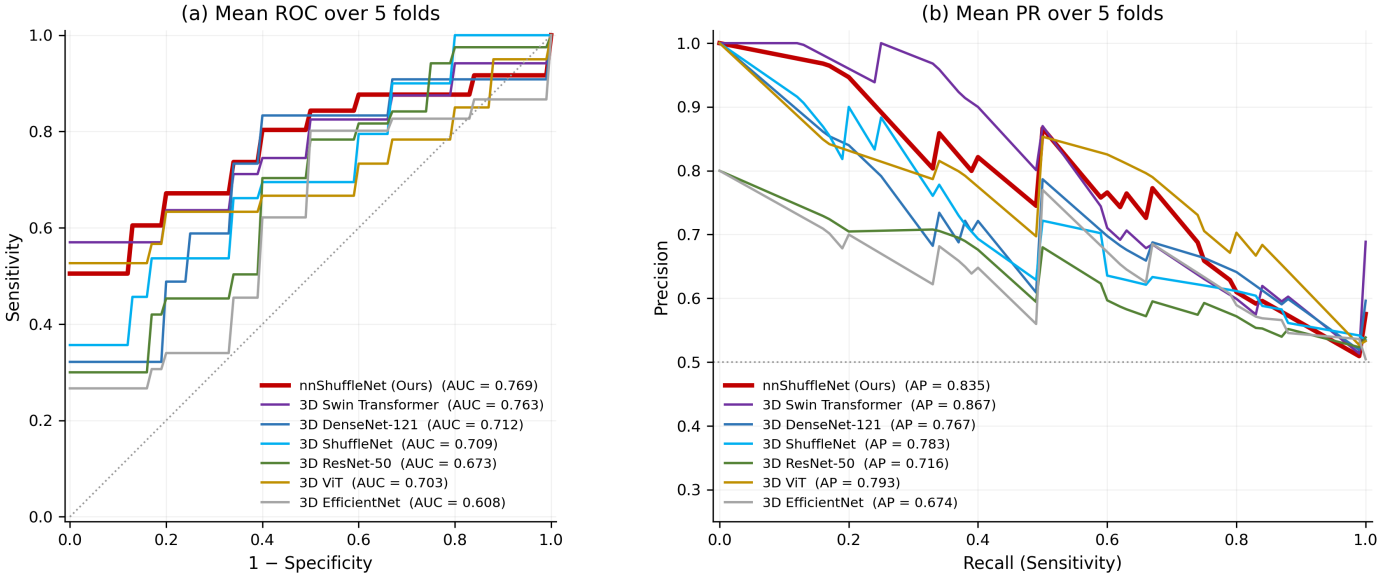


Figure 3. Patient-level ROC and PR curves under patient-exclusive five-fold cross-validation. (**a**) Mean ROC curves, with legend values giving the arithmetic mean of five fold AUCs. (**b**) Mean PR curves, with legend values giving the arithmetic mean of five fold AP values. The supplied PR panel uses non-HCC as the positive class; the ROC AUC is numerically unchanged when the binary class and score are jointly complemented. All curves use patient-level test predictions, and no 95% confidence band is shown.

We present the patient-level prediction scatter for the same seven models in Figure 4. Each point represents one independent patient, giving 54 points per model (27 HCC and 27 non-HCC) across the five outer test folds. Patients are arranged by model score within each fold on the horizontal axis; the vertical axis shows the patient-level non-HCC probability averaged over that patient's volumes, which is the complement of the HCC probability used for the main patient-level evaluation. The points are predictions rather than PCA features or a two-dimensional feature embedding. The dashed line marks the plotted 0.5 decision threshold, and outlined points indicate threshold-based errors in the supplied visualization.

3.2. Volume-Level Comparisons on Unseen Patients

We report the secondary volume-level results from the same patient-exclusive outer test folds in Table 4. Our proposed nnShuffleNet achieved a mean volume-level AUC of 0.748 ± 0.115. We note that the patient-level AUC can exceed the volume-level AUC because averaging multiple predictions from one patient reduces volume-specific variation and assigns equal weight to every patient, whereas the volume analysis gives greater weight to patients contributing more volumes.

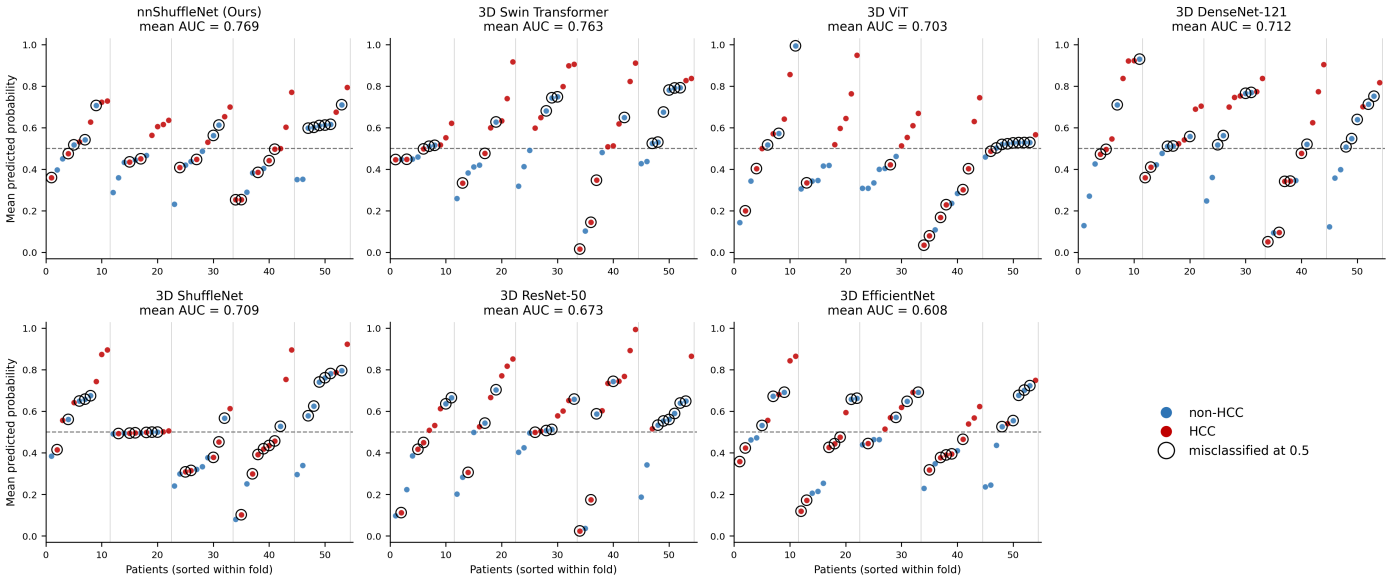


Figure 4. Patient-level prediction scatter for seven models across the patient-exclusive outer test folds. Each panel contains one point for each of the 54 patients (27 HCC and 27 non-HCC). Points are ordered by their plotted non-HCC probability within each fold; vertical partitions distinguish folds. The dashed line denotes a 0.5 threshold, and outlined points indicate threshold-based errors. The plotted quantity is an aggregated model prediction, not a learned feature or image volume.

3.3. Model Complexity and Deployment Efficiency

To evaluate structural cost and offline model-forward efficiency, we measured archived fold-1 checkpoints on the same NVIDIA RTX A6000. Input tensors were preloaded; therefore, our measurements exclude CT loading, preprocessing, transfer, visualization, and clinical-workflow time. We compare our nnShuffleNet with our 3D ResNet-50 implementation in Table 5. nnShuffleNet used 33.022M parameters and 2.768G counted Conv3d/Linear MACs, corresponding to reductions of 28.46% and 45.37%, respectively. Its measured P50/P95 forward-pass latency was 27.557/27.797 ms, and peak GPU memory was 171.19 MiB. We note that these values characterize the tested offline inference setting and are not measurements of end-to-end or real-time clinical operation.

Table 5. Offline model-forward cost comparison between nnShuffleNet and our 3D ResNet-50 implementation. Lower values are better; bold indicates the lowest value in each column. M denotes million parameters; ↓ indicates lower-is-better.

Configuration	Params M ↓	MACs G ↓	P50 (ms) ↓	P95 (ms) ↓	GPU Mem (MiB) ↓
nnShuffleNet	**33.022**	**2.768**	**27.557**	**27.797**	**171.19**
3D ResNet-50	46.159	5.067	31.944	32.212	228.32

4. Discussion

4.1. Main Empirical Findings

Our primary finding is that our proposed nnShuffleNet achieved a mean patient-level AUC of 0.769 under patient-exclusive five-fold validation. This result is based on 54 outer-test patient predictions, each obtained from a model that did not use that patient's volumes for training or validation. On the same degrouped backbone, we observed that replacing FC with PCA50-OLS increased the mean patient AUC of our nnShuffleNet from 0.719 to 0.769. The corresponding volume-level AUC was 0.748, which we use as a secondary comparison at the individual-volume evaluation unit.

4.2. Coordination of Degrouped Representation and PCA-OLS Classification Head

Degrouped convolution increases channel connections between spatial features, while PCA50-OLS reconstructs the classification mapping after backbone training through dimensionality reduction and closed-form linear fitting. Our same-backbone comparison isolates the classification-head change while retaining the learned degrouped representation. In this cohort, we observed that their combination produced higher mean patient- and volume-level AUCs than the corresponding degrouped FC configuration.

From a mathematical perspective, PBDL is a principal-component-regression-style classifier fitted on frozen deep features; PCA, OLS, and explicit intercepts are established techniques [24], and the workflow is related to fitting linear classifiers after freezing representations [25]. We fix 50 components because this provides a common low-dimensional representation across folds, keeps the OLS system overdetermined, and avoids a ridge-penalty search.

4.3. Patient-Level Aggregation

The patient-level score averages all available volume probabilities from one patient. This aggregation reduces the influence of a single atypical acquisition and gives every patient equal weight. By contrast, the volume-level analysis weights patients in proportion to their number of volumes. We attribute the difference between the patient-level mean AUC (0.769) and the volume-level mean AUC (0.748) to these different evaluation units, without implying patient overlap between training and testing.

4.4. Structural Cost and Deployment Implementation

Relative to our 3D ResNet-50 implementation, nnShuffleNet reduced the parameter count and counted Conv3d/Linear MACs by 28.46% and 45.37%. After fitting, PCA centering, projection, and the OLS mapping can be algebraically fused into one affine operation. Together with the measured latency and memory values, we consider this to provide an auditable offline implementation reference for the combined model.

4.5. Limitations and Future Research

We acknowledge several limitations in this study. First, this is a single-institution study of 54 patients, and the patient-bootstrap confidence intervals remain wide; independent multi-center evaluation is required. Second, the present task is binary HCC versus a heterogeneous non-HCC category and does not address subtype or stage classification. Third, the paired AUC differences did not remain statistically significant after Holm correction; therefore, the observed AUC rankings are interpreted descriptively within this cohort. Finally, the latency and memory measurements represent preloaded-tensor offline model forward passes on one GPU and do not establish end-to-end or real-time clinical performance. We plan future work to include external validation, physical-spacing-aware preprocessing, and prospective workflow evaluation.

5. Conclusions

We empirically evaluate our proposed nnShuffleNet, which combines a degrouped 3D ShuffleNet backbone with a PBDL head implemented as PCA50-OLS. Under patient-exclusive five-fold cross-validation of 54 independent patients, nnShuffleNet achieved a mean patient-level AUC of 0.769 $\pm$ 0.125 (95% CI: 0.635–0.883). The corresponding volume-level AUC on the same unseen-patient folds was 0.748 $\pm$ 0.115. Relative to the same degrouped backbone with FC, the patient-level and volume-level mean AUCs increased by 0.050 and 0.047, respectively. Compared with our 3D ResNet-50 implementation, the combined model reduced the parameter count by 28.46% and counted MACs by 45.37%. We consider these findings to establish a reproducible patient-exclusive empirical baseline for the evaluated cohort.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/electronics15235536/s1. Table S1. Anonymized patient-level fold assignments, class labels, and per-patient volume counts.

Author Contributions: Conceptualization, methodology, software, validation, formal analysis, writing—original draft preparation, and writing—review and editing: J.Z.; investigation, resources, data curation, visualization, supervision, and project administration: Y.L. 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 reviewed and approved by the Medical Ethics Committee of Guangxi Medical University (Approval No. KY20260007, Date: 4 January 2026). The review opinion was "Approved", and use of the data for this study was authorized under this approval.

Informed Consent Statement: Written informed consent was obtained from all patients included in this study.

Data Availability Statement: The raw CT data are subject to patient-privacy and ethics-approval restrictions and may be obtained by contacting the corresponding author. Anonymized patient-level fold assignments, experimental configurations, and analysis code are provided with the research materials.

Acknowledgments: The authors would like to thank City University of Macau for their support. This work was done when Jiaqi Zhao worked in City University of Macau as research assistant.

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

References

1. Oh, J.H.; Jun, D.W. The latest global burden of liver cancer: A past and present threat. *Clin. Mol. Hepatol.* **2023**, *29*, 355–357.
2. McGlynn, K.A.; Petrick, J.L.; El-Serag, H.B. Epidemiology of Hepatocellular Carcinoma. *Hepatology* **2021**, *73*, 4–13.
3. Chan, S.L.; Sun, H.C.; Xu, Y.; Zeng, H.; El-Serag, H.B.; Lee, J.M.; Schwartz, M.E.; Finn, R.S.; Seong, J.; Wang, X.W.; et al. The Lancet Commission on addressing the global hepatocellular carcinoma burden: Comprehensive strategies from prevention to treatment. *Lancet* **2025**, *406*, 731–778. https://doi.org/10.1016/S0140-6736(25)01042-6
4. Ying, H.; Liu, X.; Zhang, M.; Ren, Y.; Zhen, S.; Wang, X.; Liu, B.; Hu, P.; Duan, L.; Cai, M.; et al. A Multicenter Clinical AI System Study for Detection and Diagnosis of Focal Liver Lesions. *Nat. Commun.* **2024**, *15*, 1131.
5. Mulé, S.; Ronot, M.; Ghosn, M.; Sartoris, R.; Corrias, G.; Reizine, E.; Morard, V.; Quelever, R.; Dumont, L.; Londono, J.H.; et al. Automated CT LI-RADS v2018 Scoring of Liver Observations Using Machine Learning: A Multivendor, Multicentre Retrospective Study. *JHEP Rep.* **2023**, *5*, 100857.
6. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning spatiotemporal features with 3D convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 4489–4497.
7. 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. https://doi.org/10.3390/info16030195
8. Sarfati, E.; Bone, A.; Rohe, M.-M.; Aube, C.; Ronot, M.; Gori, P.; Bloch, I. Guiding the classification of hepatocellular carcinoma on 3D CT-scans using deep and handcrafted radiological features. In Proceedings of the IEEE 22nd International Symposium on Biomedical Imaging (ISBI), Houston, TX, USA, 14–17 April 2025; pp. 1–5.

9. Chang, S.C.; Yu, H.P.; Hsien, Y.-H.; Wang, P.; Wang, W.; Su, T.-H. ReViT: A hybrid approach for BCLC staging of hepatocellular carcinoma using 3D CT with multiple instance learning. In Proceedings of the IEEE EMBS International Conference on Biomedical and Health Informatics (BHI), Houston, TX, USA, 10–13 November 2024; pp. 1–8.
10. Parimala, A.B.; Shanmugasundaram, R. Classification of liver lesion stages using pyRadiomics features combined with 3D-CNN in 3D-CT and US images. *Indian J. Sci. Technol.* **2024**, *17*, 4296–4306.
11. Firat, H.; Hanbay, D. Classification of hyperspectral images using 3D CNN based ResNet50. In Proceedings of the 29th Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkey, 9–11 June 2021; pp. 1–4.
12. Lou, M.; Ying, H.; Liu, X.; Zhou, H.; Zhang, Y.; Yu, Y. SDR-Former: A Siamese Dual-Resolution Transformer for Liver Lesion Classification Using 3D Multi-Phase Imaging. *Neural Netw.* **2025**, *185*, 107228. https://doi.org/10.1016/j.neunet.2025.107228
13. Bône, A.; Rohé, M.M. Leveraging Multi-Phase CT for the Automatic Diagnosis of HCC with Deep Learning: Impacts of Lesion Localization and Phase Inclusion. In Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia, 18–21 April 2023; pp. 1–5. https://doi.org/10.1109/ISBI53787.2023.10230692
14. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
15. Köpüklü, O.; Kose, N.; Gunduz, A.; Rigoll, G. Resource efficient 3D convolutional networks. *arXiv* **2019**, arXiv:1904.02422.
16. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
17. Powers, D.J. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
18. Huang, J.; Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 299–310.
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
20. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
21. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
22. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; Houlsby, N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual Event, Austria, 3–7 May 2021.
23. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
24. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
25. Kornblith, S.; Shlens, J.; Le, Q.V. Do Better ImageNet Models Transfer Better? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 2661–2671.

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.