This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Multi-Source Feature Fusion for LI-RADS Classification in Multi-Phase CT

Yi Yang[1], Guillaume Tochon[1], Pierre-Henri Conze[2,3], Thierry Géraud[1], and Élodie Puybareau[1⋆]

[1] EPITA, LRE, Paris, France
elodie.puybareau@epita.fr
[2] IMT Atlantique, France
[3] LaTIM, UMR 1101, Inserm, Brest, France

Abstract. Automatic Liver Imaging Reporting and Data System (LI-RADS) classification for multi-phase CT remains a challenging task due to the complexity of the data and the diversity of the tumors. The AMPLIFAI Challenge proposes to classify individual lesions using a complex dataset composed of multi-phase CT scans with limited sample size, class imbalance, and missing phases. For this challenge, we propose a method combining image-based features with morphological ones to classify the lesions. Firstly, arterial phase (ART) and portal venous phase (VEN) CT are used as dual-channel inputs to train a 3D segmentation network with lesion-mask supervision, whose encoder extracts lesion-related visual features. Subsequently, a 3D CNN extracts lesion morphological features, while PyRadiomics [1] provides shape, intensity, and texture features; four additional DEL–VEN intensity-change features are also incorporated. Finally, the three feature types are fused for seven-class LI-RADS prediction using a category-weighted loss.
The experimental results show that the fusion of the three feature sources achieved the highest final score on the official validation set, and the CT encoder trained with lesion segmentation supervision achieved a higher validation final score than the evaluated general medical image pre-trained encoders.
As of 00:00 CEST on August 17, 2026, the proposed method achieved a score on the official validation set that exceeded all results reported on the validation leaderboard and ranked second on the provisional official test leaderboard.

Keywords: LI-RADS; Multiphase CT; Three-dimensional Convolutional Neural Network; Radiomics; Feature fusion; Category imbalance

1 Introduction

Hepatocellular carcinoma is a common primary malignant tumor of the liver, and early detection and accurate diagnosis are of great significance for treatment decision-making and patient prognosis [2]. The Liver Imaging Reporting

⋆ Corresponding author.

and Data System (LI-RADS) proposed by the American College of Radiology standardizes the evaluation of liver lesions in high-risk populations through categories such as LR-1 to LR-5, LR-M, and LR-TIV, thereby improving the consistency of imaging diagnosis [3].

LI-RADS characterizes the appearance of target lesions in different contrast-enhanced phases. The arterial phase mainly reflects early wash-in of lesions, while the portal venous phase and delayed phase are helpful in observing washout and delayed enhancement patterns [3]. Therefore, automatic LI-RADS classification requires not only to analyze the characteristics of lesions in a single phase, but also to integrate the changes between different phases.

The AMPLIFAI Challenge provides multi-phase CT, target lesion masks, and case-level LI-RADS labels, establishing a unified evaluation platform for this task [8].

However, this task still faces multiple challenges. The number of cases in different LR categories varies significantly, and the samples in low-level categories are limited, making the model prone to bias towards the majority categories. The overall size of the dataset is small, and using case-level labels directly to train a 3D classification network can easily lead to overfitting. In addition, the CT phases included in different cases are not completely consistent, which limits the model's learning of complete multi-phase changes [8].

This paper proposes a multi-source feature fusion method under these constraints. Firstly, ART and VEN are used as dual-channel inputs, and a CT encoder is trained using a lesion segmentation task to compensate for the lack of case-level labels with voxel-level supervision. Subsequently, a three-dimensional CNN and PyRadiomics [1] are used to extract lesion morphology and radiomics features, and the three types of features are fused to complete seven-class LI-RADS prediction.

Rather than introducing a new learning architecture, the main contribution of this work lies in the task-specific integration and evaluation of complementary visual, morphological, and radiomic information for LI-RADS classification under small-sample and incomplete multi-phase CT conditions. The main contributions of this paper are as follows:

1. We investigate lesion-segmentation-supervised representation learning for extracting target-lesion-related visual features from multi-phase CT.
2. We construct a LI-RADS classification framework that integrates CT visual features, 3D lesion morphology features, and compact radiomics features.
3. We compare different feature branches, visual encoders, and morphological encoders to evaluate the contribution of complementary information sources to classification performance.

2 Related Work

2.1 Automatic LI-RADS Classification

Existing automatic LI-RADS studies typically use multi-phase images as multi-channel inputs and directly predict lesion grades using CNNs or Transformers.

Yamashita et al. [4] used multi-phase CT and MRI to train CNN and verified the feasibility of transfer learning in small-sample LI-RADS classification; Xu et al. [5] used Swin Transformer to analyze multi-phase CT scans, further demonstrating that different contrast-enhanced phases have complementary value. However, end-to-end classification models are still susceptible to the effects of insufficient samples, imbalanced categories, and missing phases.

2.2 Learning of 3D Medical Image Representation

3D CNNs can directly process volumetric data and are widely used in medical image segmentation and classification. MedicalNet [6] provides 3D ResNet models pre-trained on medical imaging datasets, whose weights can be transferred to downstream 3D medical image tasks, while VoCo [7] learns 3D spatial representations through self-supervised learning. These methods can reduce the training difficulty of downstream tasks, but there may still be a gap between general pre-training objectives and specific lesion classification tasks. nnU-Net [9] can automatically configure the medical image segmentation process based on data characteristics. This paper uses the ResidualEncoderUNet architecture of nnU-Net for target lesion segmentation and transfers its encoder to case-level classification.

2.3 Radiomics

Radiomics converts image data into quantitative features by calculating shape, intensity, and texture features from medical images and regions of interest [10]. PyRadiomics provides an open-source platform for radiomic feature extraction [1]. Compared with deep features, radiomics have clearer statistical interpretability, but an excessive number of features may also increase the risk of overfitting in small-sample models. To this end, this paper uses a compact radiomic representation and incorporates local contrast changes between DEL and VEN, which are then fused with CT visual features and lesion morphology features.

3 Methods

3.1 Overall Framework

The proposed framework integrates CT visual, lesion morphology, and radiomic features for seven-class LI-RADS prediction, as illustrated in Fig 1.

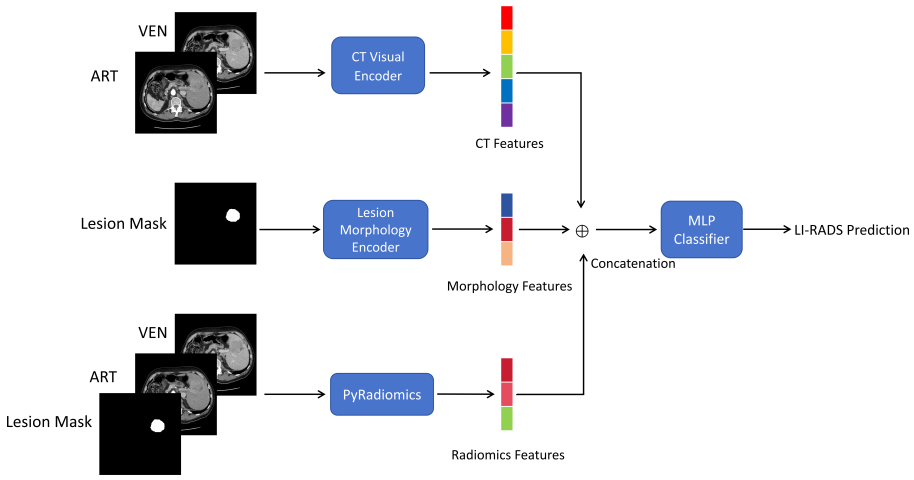

Fig. 1. Overall framework of the proposed LI-RADS classification method.

3.2 Spatial Standardization of the Lesion Center

Notable variations were observed across cases in terms of scan range, image size, and voxel spacing; consequently, the direct utilization of the full CT scan may introduce a significant amount of information unrelated to the target lesion. To mitigate these variations, we centered the crop on the lesion and extracted a three-dimensional region with the same physical extent from each case. Specifically, the crop center along each spatial dimension was determined from the bounding box of the binary lesion mask as

$$c_i = \frac{p_i^{\min} + p_i^{\max}}{2},$$

where $p_i^{\min}$ and $p_i^{\max}$ denote the minimum and maximum lesion coordinates along dimension i. A cubic region with a physical field of view of $204.8 \times 204.8 \times 204.8$ mm^3 was then cropped around this center.

The cropped multi-phase CT images and lesion masks were resampled to $256 \times 256 \times 256$ voxels with an isotropic spacing of 0.8 mm. Trilinear interpolation was used for CT images and nearest-neighbor interpolation for lesion masks. Following this step, the physical extent, voxel spacing, and array size were consistent across different cases.

3.3 Visual Feature Learning from Multi-phase CT Images

To extract visual features related to LI-RADS classification from multi-phase CT images, three-dimensional CNNs were employed as feature extractors. As

the number of available cases was limited, training a classification network from scratch using case-level LR labels alone could not provide sufficient supervision for the encoder. Therefore, ART and VEN were combined into a dual-channel input, and a 3D ResidualEncoderUNet with the ResEnc L configuration [11] was trained for lesion segmentation using lesion masks as the supervision signal.

Upon completion of the segmentation training, the network decoder was removed and the encoder parameters were frozen. ART and VEN images were fed into the encoder, and the deepest 320-channel feature map underwent 3D global average pooling to obtain a 320-dimensional case-level CT visual representation:

$$\mathbf{h}_{\mathrm{CT}} = \mathrm{Pool}\left(E_{\mathrm{CT}}\left([X_{\mathrm{ART}} \parallel X_{\mathrm{VEN}}]\right)\right) \tag{1}$$

Here, X_{ART} and X_{VEN} represent the preprocessed ART and VEN images, respectively; $\parallel$ denotes channel concatenation; E_{CT} denotes the encoder trained for segmentation; Pool denotes 3D global average pooling; and $\mathbf{h}_{\mathrm{CT}}$ denotes the final case-level CT visual features. As opposed to training directly on a classification task, segmentation supervision is intended to guide the encoder to focus on the target lesion area and reduce the influence of surrounding irrelevant tissue on the visual representation, while providing voxel-level supervision signals when case-level labels are limited.

3.4 Lesion Morphology and Radiomic Characterization

The lesion mask directly describes the size, contour, and overall structure of the lesion in three-dimensional space. To this end, we used a lightweight 3D CNN to process the binary lesion mask directly. The network was first trained independently for LI-RADS classification; after training, its classification head was removed, and a frozen encoder was used to generate 64-dimensional lesion morphological features.

Meanwhile, we used PyRadiomics [1] to extract a 66-dimensional radiomic representation, consisting of 14 three-dimensional shape features, 18 ART first-order features, 24 ART GLCM features, 5 ART NGTDM features, and 5 VEN NGTDM features.

Given that LI-RADS classification also relies on changes between different enhancement phases [3], we further incorporated four DEL–VEN intensity-change features from the inner lesion boundary. The inner boundary was defined as lesion voxels located within 2 mm of the lesion contour. VEN and DEL images were resampled onto the same spatial grid before computing the intensity differences, but no explicit inter-phase registration was performed. Therefore, inter-phase misalignment caused by respiratory motion or patient movement may affect the DEL–VEN intensity-change features to some extent

For each voxel in this region, the intensity difference was calculated as

$$\Delta I(v) = I_{\mathrm{DEL}}(v) - I_{\mathrm{VEN}}(v). \tag{2}$$

The mean, 10th percentile, median, and 90th percentile of ΔI were used as four additional features, resulting in a final 70-dimensional radiomic representation. For cases without a DEL phase, these four features were treated as missing

values and imputed using the corresponding median values estimated from the training data.

The lesion morphology branch describes the three-dimensional geometric structure of the lesion through a learning-based approach, while the radiomics branch provides explicit information on shape, texture, and enhancement changes. Together, they complement information that the CT visual encoder might overlook.

3.5 Feature Fusion and Classification

For each case, we concatenated the outputs from the three branches into a unified case-level representation:

$$\mathbf{h} = \mathbf{h}_{\mathrm{CT}} \parallel \mathbf{h}_{\mathrm{shape}} \parallel \mathbf{h}_{\mathrm{rad}},$$

Here, $\mathbf{h}_{\mathrm{CT}}$ represents the CT visual features extracted by the CT encoder; $\mathbf{h}_{\mathrm{shape}}$ represents the morphological features extracted by the lesion mask encoder; $\mathbf{h}_{\mathrm{rad}}$ represents the radiomic features; the symbol $\parallel$ denotes feature concatenation; and $\mathbf{h}$ denotes the final fused features.

The fused features were processed through a lightweight multi-layer perceptron, which output prediction scores for the seven LI-RADS categories. The category with the highest score from the classifier was ultimately selected as the prediction result. Accounting for the significant disparity in sample sizes across different categories, we employed category-weighted cross-entropy during training to enhance the influence of less frequent categories on model optimization.

3.6 External Resources

The final submitted model was trained exclusively on the multi-phase CT scans, lesion masks, and LI-RADS labels provided by the AMPLIFAI Challenge. No external medical datasets, external pre-trained weights, or additional manual annotations were used in the final model. External pre-trained models were used only for comparison in the ablation experiments.

4 Experiments

4.1 Dataset and Evaluation Metrics

This study uses multi-phase liver CT data provided by the AMPLIFAI Challenge. The public development data comprise 531 training cases and 59 validation cases, while final challenge performance was evaluated on the private test set [8].

This dataset includes cases from LR-1 to LR-5, as well as LR-M and LR-TIV. For each case, a lesion mask is provided with scans of at least one phase. For the training phase of the challenge, some additional data were provided, such as capsule masks, but were not given for the validation and testing phases. The

dataset is highly imbalanced, with LR-5, LR-M, and LR-TIV constituting the majority of the cases, while LR-1, LR-2, and LR-3 contain only a very limited number of samples. The limited overall dataset size and severe class imbalance increase the difficulty of learning reliable representations, particularly for the low-level LR categories.

This study adopts the official evaluation metrics of the challenge. The final score is defined as

$$S_{\text{final}} = 0.85S_{\text{QWK}} + 0.15S_{\text{special}}. \tag{3}$$

where S_{QWK} denotes the Adjusted Quadratic Weighted Kappa computed for cases whose true and predicted labels are both within LR-1 to LR-5, with negative values clipped to zero. S_{special} denotes the balanced accuracy for distinguishing three groups: LR-1–LR-5, LR-M, and LR-TIV. Cases with a true LR-1–LR-5 label but predicted as LR-M or LR-TIV are excluded from the QWK calculation and are penalized through S_{special}. In addition, the overall prediction accuracy across all cases is reported in this study.

4.2 Implementation Details

The segmentation network was trained for 250 epochs using 192^3 patches with a combination of Dice and cross-entropy losses. The final classifier consisted of two fully connected layers, with ReLU activation and dropout applied between the two layers. It was trained using category-weighted cross-entropy loss, where the weight of each category was determined by the inverse square root of its sample frequency. AdamW [12] was used with an initial learning rate of 1×10^{-3}, a weight decay of 1×10^{-4}, and a batch size of 32.

All models submitted to the hidden test set were trained using 527 of the 531 cases in the official training set, with four cases without a VEN phase excluded because the CT visual branch required paired ART and VEN inputs. The 59-case official validation set was used for model selection following the challenge protocol, while the hidden test set was used for the final challenge evaluation.

Before transfer to the LI-RADS classification task, the ART–VEN segmentation model was evaluated on the 59-case validation set using whole-volume inference, achieving a foreground lesion Dice of 0.6489.

4.3 Classification Results of the Complete Method

We report the performance of the proposed method on the official validation set used during model development and on the hidden test set used for the challenge evaluation. The results are presented in Table 1.

Table 1. Overall Results of the Complete Method

Evaluation Data	**Accuracy**	**Adjusted QWK**	**Special Category Recognition**	**Final Score**
Official Validation Set	62.71%	0.7706	0.6942	0.7591
Official Test Set	—	0.42	0.47	**0.43**

As of 00:00 CEST on August 17, 2026, the final score achieved by the proposed method on the official validation set exceeded all results reported on the validation leaderboard [13]. At the same time, our submission ranked second on the provisional official test leaderboard, with a final score of 0.43 [13]

The model achieved an Adjusted QWK of 0.7706 on the official validation set, indicating relatively good ordered consistency between the predictions and the true LR levels on this validation set.

A substantial performance gap was observed between the official validation set and the hidden test set. A comparable validation-to-test performance decrease was also observed for multiple submissions on the challenge leaderboard, indicating that this discrepancy was not specific to our method [13]

4.4 Ablation Study

Impact of Feature Branches To investigate the contribution of each information source to classification performance, we performed ablation experiments on the complete method by removing the CT visual branch, the lesion morphology branch, and the radiomics branch, respectively. For all the experiments, the remaining feature branches and training settings were kept unchanged, while the input dimension of the classifier was adjusted according to the retained feature composition. The experimental results are displayed in Table 2.

Table 2. Feature Branch Ablation in the Complete Classification Pipeline

Feature Composition	Accuracy	Adjusted QWK	Special Category Recognition	Final Score
CT Features + Morphology Features + Radiomics Features	62.71%	**0.7706**	0.6942	**0.7591**
Morphology Features + Radiomics Features	62.71%	0.7496	**0.7180**	0.7448
CT Features + Radiomics Features	**66.10%**	0.5947	0.7093	0.6119
CT Features + Morphology Features	55.93%	0.0000	0.6028	0.0904
Radiomics Features	64.41%	0.5407	0.7243	0.5683

The radiomics branch makes the largest contribution to the final performance. After removing the radiomics branch, the accuracy, Adjusted QWK, and final score decrease substantially. This indicates that the radiomics branch provides substantial information for determining LR levels and helps preserve the ordered consistency of the predictions.

After removing the lesion morphology branch, both Adjusted QWK and final score are substantially reduced. This suggests that lesion morphology features contribute to the ordinal consistency of the predictions.

The removal of the CT visual branch results in a slight decrease in Adjusted QWK and final score. This suggests that the current CT encoder provides a positive contribution to the full model, although its contribution is smaller than that of the radiomics branch, implying that the CT visual representation could be improved. When only radiomics features were used for classification, both the Adjusted QWK and the final score decreased to some extent compared with the complete model. Meanwhile, compared with the model combining radiomics

features with the learned morphology representation, the radiomics-only model also showed lower Adjusted QWK and final score. This suggests that the learned morphology representation can complement the characterization of lesion morphology provided by handcrafted radiomic shape descriptors.

CT Visual Encoder To find a more appropriate CT visual representation for the current task, we replaced the CT encoder in the complete classification pipeline while keeping the other feature branches and training settings unchanged.Both MedicalNet ResNet10 [6] and VoCo [7] encoders are pre-trained on large-scale medical imaging datasets and were used only as replacement encoders.

Table 3. Comparison of Different CT Encoders in the Complete Classification Pipeline

CT Visual Encoder	Accuracy	Adjusted QWK	Special Category Recognition	Final Score
ResidualEncoderUNet Encoder	62.71%	**0.7706**	0.6942	**0.7591**
MedicalNet ResNet10	55.93%	0.6623	0.6554	0.6612
VoCo Encoder	**64.41%**	0.5467	**0.7005**	0.5698

With the fusion framework kept unchanged, the ResidualEncoderUNet encoder trained on lesion segmentation performed the best. This suggests that, under the evaluated setting, this encoder provided a more suitable visual representation for LI-RADS classification than the evaluated general pre-trained encoders.

Lesion Morphology Encoder To investigate the capability of different encoders for characterizing lesion morphology, we employed a 3D CNN, a 3D residual CNN, and Tiny VoxNet [14] to process binary lesion masks. The experiment only replaced the lesion morphology encoder in the complete classification process, while keeping the other feature branches, classifier, and training settings unchanged

Table 4. Comparison of Different Lesion Morphology Encoders in the Complete Classification Pipeline

Lesion Morphology Encoder	Accuracy	Adjusted QWK	Special Category Recognition	Final Score
3D CNN	**62.71%**	**0.7706**	**0.6942**	**0.7591**
3D Residual CNN	54.24%	0.4358	0.6679	0.4706
Tiny VoxNet	49.15%	0.4087	0.6378	0.4431

The results show that the 3D CNN achieved the highest final score, whereas the 3D residual CNN and Tiny VoxNet did not yield performance improvements. These results suggest that the relatively simple 3D CNN was more suitable for

the current small-sample setting. Therefore, we adopted it as the lesion morphology encoder.

4.5 Discussion

Our experimental results show that the fusion of CT visual features, lesion morphology features, and radiomics features achieved the highest final score on the official validation set. However, there are still three limitations of the current method. The first is the severe class imbalance of the training set, which contains very few LR-1 to LR-3 cases, so that the model cannot acquire sufficiently stable representations for low-level categories. The second limitation is that some training cases do not contain complete multi-phase CT scans [8], thus limiting the model's ability to extract information from each contrast-enhanced phase and their dynamic changes. A third limitation is the small overall dataset [8], which increases the risk of model overfitting and may limit the reliability of validation-based model selection.

A substantial validation-to-test performance decrease was also observed for multiple submissions on the challenge leaderboard [13], suggesting that the relatively small official validation set may provide only a limited estimate of hidden-test performance.

5 Conclusion

This paper presents a LI-RADS classification method for multi-phase CT that integrates image-based, 3D lesion morphology-based, and radiomic-based features. The proposed method employs a segmentation-trained CT encoder, a 3D CNN for lesion morphology, and a compact radiomic representation, followed by seven-class prediction using a category-weighted loss. On the official validation set, the three-source fusion achieved a higher final score than the evaluated branch-removal variants, while the ResidualEncoderUNet encoder achieved a higher validation final score than the evaluated general medical image pretrained encoders.

As of 00:00 CEST on August 17, 2026, our method exceeded all results reported on the validation leaderboard and ranked second on the provisional official test leaderboard [13]. Future work will focus on missing and multi-phase information under small-sample conditions and improving performance on lower-grade LI-RADS cases.

Disclosure of Interests

The authors declare no competing interests relevant to this work.

References

1. van Griethuysen, J.J.M., Fedorov, A., Parmar, C., et al.: Computational radiomics system to decode the radiographic phenotype. Cancer Research 77(21), e104–e107 (2017)
2. Llovet, J.M., Kelley, R.K., Villanueva, A., et al.: Hepatocellular carcinoma. Nature Reviews Disease Primers 7, 6 (2021)
3. Chernyak, V., Fowler, K.J., Kamaya, A., et al.: Liver Imaging Reporting and Data System (LI-RADS) Version 2018: Imaging of hepatocellular carcinoma in at-risk patients. Radiology 289(3), 816–830 (2018)
4. Yamashita, R., Mittendorf, A., Zhu, Z., Fowler, K.J., Santillan, C.S., Sirlin, C.B., Bashir, M.R., Do, R.K.G.: Deep convolutional neural network applied to the Liver Imaging Reporting and Data System (LI-RADS) version 2014 category classification: a pilot study. Abdominal Radiology 45, 24–35 (2020)
5. Xu, Y., Zhou, C., He, X., Song, R., Liu, Y., Zhang, H., et al.: Deep learning-assisted LI-RADS grading and distinguishing hepatocellular carcinoma (HCC) from non-HCC based on multiphase CT: a two-center study. European Radiology 33, 8879–8888 (2023)
6. Chen, S., Ma, K., Zheng, Y.: Med3D: Transfer Learning for 3D Medical Image Analysis. arXiv preprint arXiv:1904.00625 (2019)
7. Wu, L., Zhuang, J., Chen, H.: VoCo: A Simple-yet-Effective Volume Contrastive Learning Framework for 3D Medical Image Analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22873–22882 (2024)
8. Kulkarni, P., Shah, N., Suryavanshi, A., Delfino, J., Tonascia, J., Wong-You-Cheong, J., Lane, B., Chirico, J., Hirsch, J.D., Li, A., Huang, H., Doo, F.X.: AMPLIFAI: A Multiphase CT Dataset for Benchmarking Clinical Reasoning in LI-RADS Assessment of Liver Lesions. Preprint (2026)
9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18, 203–211 (2021)
10. Lambin, P., Rios-Velazquez, E., Leijenaar, R., et al.: Radiomics: Extracting more information from medical images using advanced feature analysis. European Journal of Cancer 48(4), 441–446 (2012)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation. arXiv preprint arXiv:2404.09556 (2024)
12. Loshchilov, I., Hutter, F.: Decoupled Weight Decay Regularization. In: International Conference on Learning Representations (ICLR) (2019)
13. AMPLIFAI Challenge: Official Codabench leaderboard. Codabench (accessed 17 August 2026)
14. Maturana, D., Scherer, S.: VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 922–928 (2015)