European Journal of Surgical Oncology xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com

# Preoperative detection of hepatocellular carcinoma's microvascular invasion on CT-scan by machine learning and radiomics: A preliminary analysis

Simone Famularo [a,b,c,*], Camilla Penzo [d], Cesare Maino [e], Flavio Milana [b,f], Riccardo Oliva [c], Jacques Marescaux [c], Michele Diana [c,g,l], Fabrizio Romano [h], Felice Giuliante [a], Francesco Ardito [a], Gian Luca Grazi [i], Matteo Donadon [j,k,1], Guido Torzilli [b,f,1]

[a] Hepatobiliary Surgery Unit, Fondazione Policlinico Universitario A. Gemelli, IRCCS, Catholic University of the Sacred Heart, Rome, Italy
[b] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy
[c] IRCAD, Research Institute Against Cancer of the Digestive System, 1 Place de l'Hôpital, Strasbourg, 67091, France
[d] Pole d'Expertise de la Regulation Numérique (PEReN), Paris, France
[e] Department of Radiology, San Gerardo Hospital, Monza, Italy
[f] Department of Hepatobiliary and General Surgery, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy
[g] Department of General, Digestive and Endocrine Surgery, University Hospital of Strasbourg, France
[h] School of Medicine and Surgery, University of Milan-Bicocca, Department of Surgery, San Gerardo Hospital, Monza, Italy
[i] Division of Hepatobiliarypancreatic Unit, IRCCS - Regina Elena National Cancer Institute, Rome, Italy
[j] Department of Health Sciences, Università del Piemonte Orientale, Novara, Italy
[k] Department of General Surgery, University Maggiore Hospital Della Carità, Novara, Italy
[l] ICube Lab, Photonics for Health, Strasbourg, France

A B S T R A C T

*Introduction:* Microvascular invasion (MVI) is the main risk factor for overall mortality and recurrence after surgery for hepatocellular carcinoma (HCC).The aim was to train machine-learning models to predict MVI on preoperative CT scan.

*Methods:* 3-phases CT scans were retrospectively collected among 4 Italian centers. DICOM files were manually segmented to detect the liver and the tumor(s). Radiomics features were extracted from the tumoral, peritumoral and healthy liver areas in each phase. Principal component analysis (PCA) was performed to reduce the dimensions of the dataset. Data were divided between training (70%) and test (30%) sets. Random-Forest (RF), fully connected MLP Artificial neural network (neuralnet) and extreme gradient boosting (XGB) models were fitted to predict MVI. Prediction accuracy was estimated in the test set.

*Results:* Between 2008 and 2022, 218 preoperative CT scans were collected. At the histological specimen, 72(33.02%) patients had MVI. First and second order radiomics features were extracted, obtaining 672 variables. PCA selected 58 dimensions explaining >95% of the variance.In the test set, the XGB model obtained Accuracy = 68.7% (Sens: 38.1%, Spec: 83.7%, PPV: 53.3% and NPV: 73.4%). The neuralnet showed an Accuracy = 50% (Sens: 52.3%, Spec: 48.8%, PPV: 33.3%, NPV: 67.7%). RF was the best performer (Acc = 96.8%, 95%CI: 0.91–0.99, Sens: 95.2%, Spec: 97.6%, PPV: 95.2% and NPV: 97.6%).

*Conclusion:* Our model allowed a high prediction accuracy of the presence of MVI at the time of HCC diagnosis. This could lead to change the treatment allocation, the surgical extension and the follow-up strategy for those patients.

## 1. Introduction

Hepatocellular carcinoma (HCC) is the most common primary liver tumor with an increasing incidence worldwide [1]. Liver transplantation and hepatic resection represent potentially curative options [2,3]. However, relapse after surgery hinders the chance to reach the cure, with a 5-year disease recurrence up to 60% and 15% following hepatectomy or liver transplant, respectively [4–7]. Consequently, survival diminishes after recurrence to approximately 35% at 5-year after surgery [8].

Stratifying the population to identify those patients at a higher risk of recurrence remains crucial in the era of personalized medicine. With this objective, numerous prognostic factors have been emphasized, many of these histopathological. Notably, vascular invasion emerges as one of

* Corresponding author. Hepatobiliary Surgery Unit, Fondazione Policlinico Universitario A. Gemelli, IRCCS, Via Agostino Gemelli, 00100 Rome, Italy
*E-mail address:* simone.famularo@gmail.com (S. Famularo).
[1] Those authors shared the co-last authorship.

https://doi.org/10.1016/j.ejso.2024.108274
Received 9 December 2023; Received in revised form 20 February 2024; Accepted 16 March 2024
Available online 24 March 2024

the most influential elements in defining early recurrence (within <24 months) based on multiple retrospective studies [9,10]. While macrovascular tumoral invasion is easily detectable through diagnostic imaging, the microvascular counterpart, namely the presence of cancer cell clusters within a vascular space lined by endothelial cells [11], is accurately identifiable only post-operatively. Consequently, despite microvascular invasion (MVI) being reported as one of the most critical determinants of early recurrence (up to 4.4-fold increased risk [12]), its clinical utility in HCC treatment algorithm is limited. Indeed, its incidence has been reported ranging between 15% and 57.1% due to the heterogeneity of diagnostic criteria [13], yet both preoperative imaging and tumor biopsy have proven unreliable in assessing its presence [14,15].

Various algorithms have been suggested to estimate the likelihood of MVI at HCC diagnosis, linking its occurrence to factors such as tumor size, number of tumors, or AFP/PIVKA-II level [12,16–18]. However, these proposals have yielded controversial outcomes. This performance disparity has paved the way for the prognostication branch of artificial intelligence [19]. Indeed, the potential to model patient characteristics seems to improve the accuracy of risk stratifying patients in comparison to conventional models [20,21]. Notably, the analysis of preoperative imaging, involving the extraction and interpretation of radiomics features, has demonstrated the most promising outcomes [22]. Nevertheless, numerous studies have documented experiences where radiomics features and clinical factors were combined in order to optimize accuracy in the designated model [17,18,23].

The aim of this study was to investigate whether radiomics features from preoperative contrast-enhanced computed tomography (CT) scans of patients undergoing liver resection for HCC could unveil the presence of MVI in a prediction model based exclusively on imaging analysis.

## 2. Methods

### 2.1. Register information, study overview, patient selection and study design

This retrospective study evaluated the clinical data from patients prospectively enrolled by the Italian national register on hepatocellular carcinoma: the HE.RC.O.LE.S. (Hepatocarcinoma Recurrence on the Liver Study) group. This register collected data between 2008 and 2023, and 33 surgical centers are participating.The study protocols followed the ethical guidelines of the 1975 Declaration of Helsinki (as revised in Brazil 2013) [24]. The Ethical Committee of the coordinating center (San Gerardo Hospital, Monza, Italy, "Monza e Brianza Ethical Committee") reviewed and approved the HE.RC.O.LE.S. protocol on December 21, 2018. Surgery data were previously described [4]. The indication for surgery was assessed by multi-disciplinary meetings patient by patient - involving surgeons, hepatologist, oncologist, radiologist, interventional radiologist, infectologist - as the sum of different evaluations about underlying liver function, tumor burden and comorbidities, creating tailored treatment for each single case. Among the participating centers, a new protocol dedicated to the present study was setted up (ART.I.HCC. Study, clinicaltrial.gov: NCT05637788). Fifteen centers approved the protocol. For the preliminary results presented here, data from 4 centers were analyzed.

Among patients already enrolled in the HE.RC.O.LE.S. register, it was required to provide the anonymized triphasic CT scan collected maximum 30 days before surgery.

All data were submitted by local researchers and anonymized before the submission to the coordinating center. Data collection was done using an electronic database system in all centers. Thus, the DICOM file for each CT scan was stored in an online cloud (Sync.com) in compliance with the European privacy Statement 679/2016/UE.

Results are reported according to principles of Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) [25]. All consecutive adult patients (age ≥18 years) with a radiological and/or histological proven of HCC who underwent surgery from January 2008 to December 2022 were evaluated. Inclusion criteria to be considered for this study were: 1) patients who underwent surgery for the first diagnosis of HCC without previous therapies and 2) a triphasic CT scan available in the DICOM format, acquired maximum 30 days before surgery. Exclusion criteria were: (1) a CT scan without all the three phases available, (2) missing data on the study endpoints, (3) Histopathological specimen of combined liver primary neoplasms (e.g. 'hepato-cholangiocarcinoma'), (4) Surgery as a downstaging therapy for transplant, (5) Patients treated with surgery in case of not-curative intent (palliation, best supportive care, etc).

Patients recruited in the ART.I.HCC. protocol were anonymized with an alphanumeric code, the same already assigned to the specific patient in the HE.RC.O.LE.S. register.

### 2.2. Study aims and end-point definition

The aim of the study was to develop a machine learning model to predict at the preoperative CT scan those patients who will have a MVI at the specimen. For this preliminary analysis, to assess the best predictive model, the accuracy of each model was used as the primary metric.

The presence of MVI was assessed on the histological specimen after liver resection, by dedicated pathologists in each participating center. The information was then classified as a two-levels categorical variable (0 or 1, namely "no" or "yes") into the dataset as the target variable.

### 2.3. CT scan segmentation and radiomics features extraction

The DICOM files were transferred onto a dedicated freeware statistical-based texture analysis software (LIFEx software, www.lifexsoft.org) [26]. The segmentation for texture analysis has been done independently by two radiologists.

For each tumor, the slice where the tumor was most heterogeneous were identified on delayed phase images. On this slice, a two-dimension (2D) region of interest (ROI) was manually drawn, following tumor margins and encompassing the entire tumor. Another region was designed around this first ROI enlarging the tumor's boundaries of 1 cm, in order to segment the peritumoral area. Finally, a ROI was designed in a healthy portion of the liver, in the opposite liver lobe. The ROIs were recorded and transposed identically on the same slices of portal-venous, arterial and unenhanced images. Finally, three ROIs were segmented for each phase: tumoral, peritumoral and healthy liver areas.

For patients with several nodules, texture analysis was performed only on the largest one, as for the morphological analysis.

The LIFEx's statistical-based model provided a set of first-order statistics to evaluate the grey-level frequency distribution from the pixel intensity histogram in a given area of interest, [including mean intensity, threshold (percentage of pixels within a specified range), entropy (irregularity), standard deviation, skewness (asymmetry), and kurtosis (peakiness or flatness histogram)], and a set of second-order statistics, accounting for the location of the pixels and their spatial relationship (including second-order entropy, energy, homogeneity, dissimilarity, and correlation). The feature extraction was made by one of the radiologists (CM). These features were extracted per each ROI and per each phase of the CT scan, collecting a total of 667 radiomics features per each patient. The complete list of the radiomics features considered was reported in s-Table 1.

### 2.4. Exploratory data analysis and features engineering

Exploratory data analysis was run to understand the data structures, and to highlight the presence of linear dependencies or collinearities: all the radiomics features with these characteristics were excluded. Since the relatively limited amount of training samples, each with a large number of features, we should only select the most important features for training so that our model doesn't need to learn for so many features

**Table 1**
Baseline characteristics of the whole cohort.

| n | Overall |
|---|---|
| | 218 |
| **Age, years (median [IQR])** | 72.00 [66.00, 76.00] |
| **Sex - Female (%)** | 44 (20.8) |
| **ECOG PS (%)** | |
| **1** | 38 (23.6) |
| **2** | 13 (8.1) |
| **Charlson Index (median [IQR])** | 6.00 [5.00, 8.00] |
| **Cirrhosis (%)** | 114 (54.3) |
| **Child-Grade B (%)** | 14 (8.1) |
| **MELD score (median [IQR])** | 7.00 [6.00, 9.00] |
| **HBV + (%)** | 39 (18.6) |
| **HCV + (%)** | 84 (40.0) |
| **Potus (%)** | 49 (23.3) |
| **Varices (%)** | 23 (11.5) |
| **Alfa-Feto-Protein (median [IQR])** | 13.00 [4.30, 127.00] |
| **N. Nodules (median [IQR])** | 1.00 [1.00, 2.00] |
| **Size, cm (median [IQR])** | 4.00 [2.50, 6.00] |
| **BCLC (%)** | |
| **0-A** | 118 (54.1) |
| **B** | 39 (17.8) |
| **C** | 61 (27.9) |
| **Macrovascular Invasion (%)** | 10 (5.0) |
| **Microvascular invasion (%)** | 72 (33.02) |

and eventually overfit. Thus, to reduce the noise created by the very high number of radiomics features extracted, a principal component analysis (PCA) was run, keeping in the new obtained dimensions the 95% variability of the original dataset.

The dataset obtained was then split in two sets by random sampling with stratification for the presence (or absence) of MVI: training (70% of the original cohort) and test (30%) sets. This also was done to reduce the risk of overfitting.

### 2.5. Models' training, tuning and testing

Three different models were fitted on the data: extreme gradient boosting (XGB), random forest (RF) and a fully connected 2-layers multilayer perceptron neural network (neuralnet).

Each model was trained in the training cohort with the task to classify the patients among those who had MVI or those who did not. To maximize the results, the hyperparameters tuning was done with the "grid search" method (namely, a process that searches exhaustively through a manually specified subset of the hyperparameter space of the targeted algorithm) with a 10-fold cross validation: the parameters' combination providing the highest accuracy in the cross-validated set was further employed for the definitive training model. The list of hyperparameters tuned, and the relative values selected for the grid search, were reported in supplementary text 1 in the supplementary materials.

After tuning, the models were trained with the selected hyperparameters in the training set. Then, the performance of the models were measured in the unseen-before test set.

### 2.6. Metrics definitions

All the predictions made by the models in the test-set were dichotomized according to a predefined cut-off of 0.5, thus confusion matrices were created to estimate the metrics. The model accuracy has been defined as the number of all correct predictions divided by the total number of the dataset. Sensibility was defined as the number of correct positive predictions divided by the total number of positives. Specificity was defined as the number of correct negative predictions divided by the total number of negatives. Positive predictive value (PPV) was defined as the ratio of the number of records, or correctly classified records, for the specified class divided by the total number of predictions for this class. The negative predictive value (NPV) was defined as the proportion of predicted negatives which are real negatives.

### 2.7. Other statistical information

Continuous variables were reported as mean (standard deviation) or median (interquartile range) as appropriate; categorical variables were reported as number and frequencies.

The problem of missing values among the independent clinical variables was tackled using Multiple Imputation by Chained Equations (MICE) method [27]. We imputed 20 datasets using predictive mean matching for continuous variables, logistic regression for binary and multinomial regression for categorical variables with more than two levels [28].

All the analyses were carried out using R (version 4.3.0; packages: mice, caret, randomForest, neuralnet, xgboost).

## 3. Results

Between 2008 and 2023, 500 patients with available triphasic CT scan in DICOM format were enrolled by the participating centers. As a preliminary analysis, 218 (39.8%) from 4 centers were analyzed according to the study protocol.

Median age was 72 years old (IQR 66–76), while 44 (20.8%) patients were female. Median Charlson Comorbidity Index was 6 (IQR 5–8). One hundred fourteen patients (54.3%) were cirrhotics, and 14 (8.1%) were Child B, while median MELD score was 7 (IQR 6–9). Thirty-nine (18.6%) patients were HBV positive, while 84 (40.0%) were HCV positive. Alcohol consumption was reported by 49 (23.3%) patients, and presence of varices was assessed in 23 (11.5%) cases. Regarding the tumor burden, median AFP was 13.0 ng/ml (IQR 4.30–127.00); median number of nodules was 1 (IQR 1–2) and median size was 4.0 cm (IQR 2.50–6.00). Considering the tumor stage, 118 (54.1%) were BCLC-0/A, 39 (17.8%) were BCLC-B and 61 patients (27.9%) were BCLC-C. Macrovascular invasion was present in 10 (5.0%) patients. At the histological specimen, presence of MVI was assessed in 72 (33.02%) patients. The baseline was reported in Table 1.

After CT scan segmentation of the tumor, peritumoral and healthy area in a single slice of each CT scan in each phase, 671 radiomics features were extracted. To reduce the noise derived by such a high number of variables, a PCA was applied, and the dataset was reduced to 58 dimensions, keeping the 95% of the variance of the original data (s-Fig. 1).

The dataset was then randomly split in two: training set (154 patients, 70%) and test set (64 patients, 30%). With the former, the hyperparameters tuning for each model was run, and the following were the best parameters able to maximize the model accuracy:

XGB model.

- N. rounds = 100
- Max. depth = 3
- Eta = 0.1
- Gamma = 0
- Col.sample by three = 0.6
- Min. child. weight = 1
- Subsample = 0.75

For the RF model.

- N. tree = 500
- Mtry = 30

For the Neuralnet model.

- N. layers = 2
- Neurons in layer 1 = 40
- Neurons in layer 2 = 20

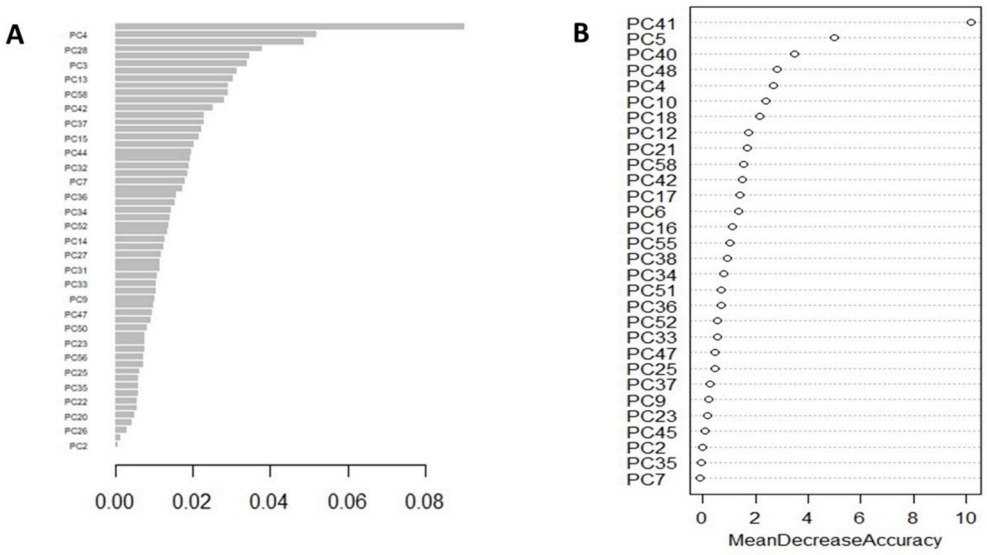


**Figure 1.** Variable importance in the (A) XGB model and (B) the Random Forest model. This figure shows how the algorithms "think", and the impact of each component in determining the risk prediction. However, after the usage of a Principal Component Analysis to reduce the dataset's noise, the clinical understanding of each component is still very difficult: this is an example of the meaning of "black box" issue in machine learning models development.

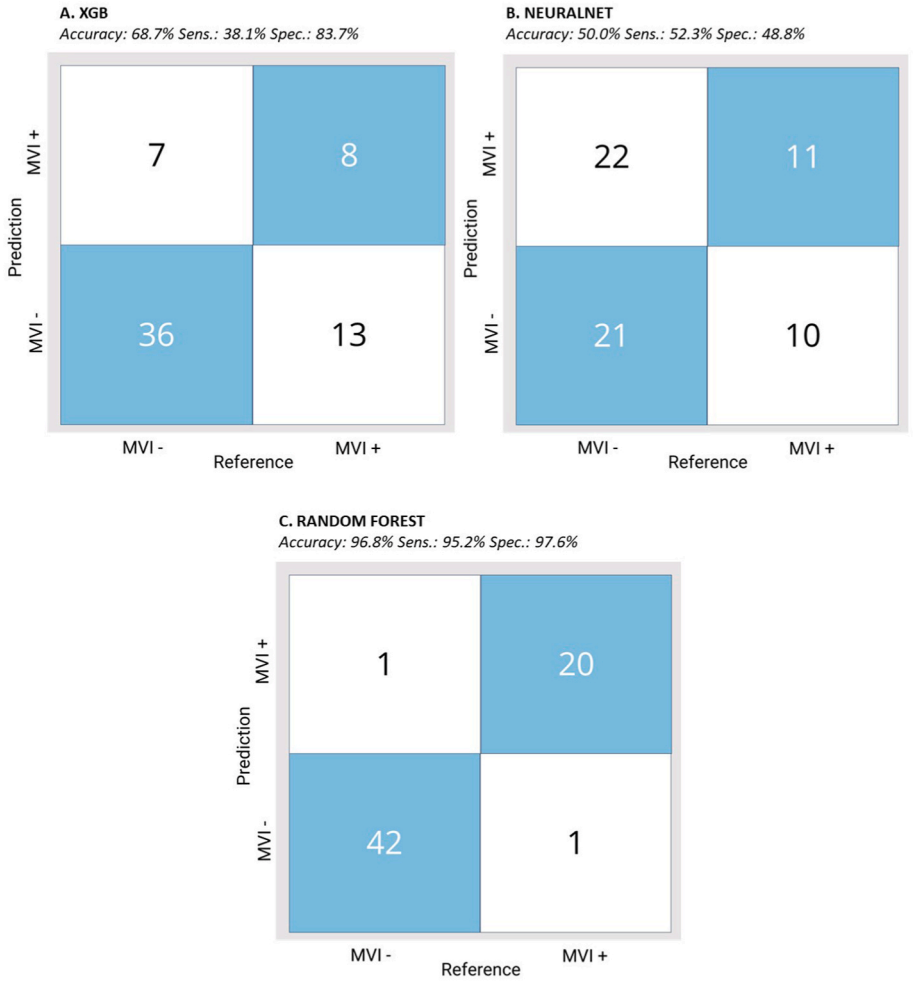


**Figure 2.** The confusion matrices reported the prediction performance of each model, estimated in the test-set. (A) XGB model, (B) Neuralnet model and (C) Random Forest model.

- Learning rate = 0.001
- Termination threshold = 1%
- Activation function = sigmoid

### 3.1. Models training results

After tuning, the models were trained with the aforementioned hyperparameters in the training cohort.

The XGB model achieved an accuracy of 98.6%, Sens.: 96.6%, Spec.:99.5%, PPV: 98.9% and NPV: 98.5%. The variable importance was depicted in Fig. 1a. The Neuralnet model showed an accuracy of 98.6%, Sens.: 98.9%, Spec.: 98.4%, PPV: 96.9% and NPV: 99.4%. The scheme of the neural network was depicted in s-Fig. 2. Finally, the RF model obtained an accuracy of 98.3%, Sens.: 94.7%, Spec.: 100%, PPV: 100% and NPV: 97.5% (the variable importance was depicted in Fig. 1b). These metrics were summarized in Table 2.

### 3.2. Models performance in the test-set

The model trained were then tested in an unseen dataset (the test-set). After a standard dichotomization of the models' prediction at 0.5 points, the confusion matrices were generated. The XGB model reached an accuracy of 68.7%, Sens.: 38.1%, Spec.: 83.7%, PPV: 53.3% and NPV: 73.4% (confusion matrix illustrated in Fig. 2a). The neuralnet model reached an accuracy of 50.0%, Sens.: 52.3%, Spec.: 48.8%, PPV: 33.3% and NPV: 67.7%, and the relative confusion matrix was depicted in Fig. 2b. Thus, the RF model showed an accuracy of 96.8%, Sens: 95.2%, Spec.: 97.6%, PPV: 95.2% and NPV: 97.6% (Fig. 2c). All the metrics in the training and the test sets were summarized in Table 2.

## 4. Discussion

The present study investigates the feasibility of constructing a predictive model to assess the MVI status using 3 different supervised machine-learning algorithms, exclusively leveraging radiomics features extracted from preoperative triphasic CT scans of patients undergoing liver resection for HCC. Our findings, based on a cohort of 218 patients from four participating centers, underscore the potential of these models as a robust tool for MVI prediction, introducing a novel perspective in the field of personalized clinical practice. In the analyzed population, MVI occurred in 33% of cases, in line with the reported literature rates [13]. Notably, the RF model demonstrated exceptional reliability in predicting MVI, achieving 96.8% accuracy in the validation cohort.

MVI not only holds significance as an independent factor for recurrence or survival, but also in potentially guiding treatment decisions at HCC diagnosis. About that, Lim et al. highlighted that HCC patients beyond the Milan criteria, in the absence of MVI, can achieve an OS comparable to those meeting the criteria [10]. Recently, it has also reported that MVI could be a sort of indirect sign of systemic disease [29]. Moreover, the presence of MVI has been shown to modify also the surgical indication, with patients submitted to anatomic resection (the theoretical gold standard for HCC) reaching similar survival of those who do not [30]. All this evidence enforced the urgency to develop a reliable instrument to predict preoperatively the presence of MVI, in order to tailor the treatments' allocation. This consideration became even more important in the era of adjuvant and neoadjuvant therapies, where recently the combination of Atezolizumab and Bevacizumab has been reported to increase the DFS in patients with worrisome features such as MVI [31]. For all these reasons, knowing MVI status at HCC diagnosis could significantly contribute to tailoring the therapeutic approach, potentially changing the actual scenario.

A peculiarity of our study is that it represents a crucial departure from existing literature by relying exclusively on radiomics features, without incorporating traditional clinical variables, which usually seems to increase the model's accuracy when combined with radiomics. In contrast to previous studies where the prediction of MVI achieved the highest accuracy by combining radiomic features with other tumor radiological characteristics (e.g., size, number of nodules, tumor capsule aspect) or clinical data (AFP or PIVKA-II level) [14,17,18,32,33] our model achieved outstanding accuracy on its own. Segal et al. [34] were among the first to associate imaging features on CT to MVI prediction, and these "worrisome" characteristics were interpreted by Renzulli et al. [17] in terms of prediction MVI accuracy in HCC. After that, the potentiality of the "hidden" data, extracted from quantitative imaging features or textures to decode tissue pathology, was analyzed. The aim was to develop predictive models based on radiomic scores in conjunction with clinical and radiological factors [18,33]. However, in these experiences, the radiomic features still resulted less relevant than radiologic scores and only hybrid models were reported to achieve the best results [32]. Our model, indeed, is one of the first obtaining high degree of performance with radiomics features only. This challenges the conventional belief that clinical variables are imperative for optimal predictive models, showcasing the potency of radiomics features alone. However, our results are preliminary, and the impact of clinical data in increasing the model's efficacy was not tested. Although the obtained metrics were already satisfactory, a comparison with mixed-models combining clinical and radiomics characteristics should be considered the natural evolution of the present study.

The data from which these results were obtained represented a large cohort, spanning from 2008 to 2022, offering a comprehensive representation of HCC cases with diverse clinical characteristics: this choice was fundamental in order to establish a model able to make the best prediction in every clinical setting. In order to reach this goal, radiomics features were extracted from 3 different liver areas that are known to play a role in the carcinogenesis of HCC: the tumor area, the peritumoral district and the non tumoral area. While the first seems quite obvious, and the latter is a consequence of the underlying liver damage that is a prerequisite for the tumorigenesis, the second was included after the interesting report of Feng et al. [35], who presented data indicating a higher incidence of MVI in the area surrounding the tumor. Moreover, the peritumoral area has been recently recognized as a significant space where the interaction between the tumor microenvironment and the immunologic host system interact, determining changement in the tumor history [36].

One concern in the field of radiomics is regarding the very high number of parameters that could be extracted, with several difficulties in determining which one could play a role in predictions.The utilization of advanced techniques such as PCA allowed to effectively handle the high dimensionality of radiomics features, reducing the dataset to 58 dimensions while retaining 95% of the original variance. This approach allowed to keep all the information provided by radiomics, however still reducing the "noise" generated by the presence of a too large amount of data. This method should be considered better rather than selecting one single feature, because this latter approach cannot guarantee the preservation of the data variance, leading to the risk of bias. However, while PCA is a strong method to reduce the dimensionality of highly complex

**Table 2**
Summary of the metrics obtained in the training and the test set per each model.

| Training set | | | | | |
|---|---|---|---|---|---|
| Model | Accuracy | Sens. | Spec. | PPV | NPV |
| XGB | 98.6 | 96.6 | 99.5 | 98.9 | 98.5 |
| NeuralNet | 98.6 | 98.9 | 98.4 | 96.9 | 99.4 |
| RF | 98.3 | 94.7 | 100 | 100 | 97.5 |
| Test Set | | | | | |
| Model | Accuracy | Sens. | Spec. | PPV | NPV |
| XGB | 68.7 | 38.1 | 83.7 | 53.3 | 73.4 |
| NeuralNet | 50 | 52.3 | 48.8 | 33.3 | 67.7 |
| RF | 96.8 | 95.2 | 97.6 | 95.2 | 97.6 |

dataset, its reproducibility and its comprehensibility in clinical terms is at least arguable. Indeed, concerns around the transparency of the model and lack of explainability (black box nature of algorithms [37]) pose challenges in the era of evidence-based medicine that relies on transparency and reproducibility in the decision-making process. It is important that these issues are better explored and addressed with a solution that should be found on robust internal and external validation.

Looking at the models tested, the Random-Forest showed the best accuracy when compared with the other models, fitting better the types of data available. This is not surprising: neural networks have a strong potentiality in increasing the efficacy of classification, but their role is better deployable in case of unsupervised learning, or with a very large amount of data. Our dataset, although it is one of the largest collecting radiomics data on HCC, was still too small to be suitable for an efficient neural network analysis.

Finally, it should be clarified that the presented results are preliminary and entail some limitations. Firstly, the CT scans were collected based on the availability of the DICOM files among the participating centers; thus, they were not necessarily consecutive, leading to the risk of selection bias. However, the percentage of MVI among this cohort was in line with the one reported in the literature and in the general dataset where the clinical data were extracted. Moreover, the problem of overfitting the model was partially mitigated by splitting the dataset randomly in two groups (one for training and one for testing), and by features' selection. However, these methodologies didn't reach the goal to avoid overfitting in the XGB and the neuralnet models. Anyway, an external validation set is mandatory to assess clearly the performance of the models even in different settings, facing the real generalizability of the presented models. Despite the relatively high number of cases available in this study considering the rest of the literature, many clinical scenarios could not be represented by the present cohort, reducing the generalizability of the models. The choice of sampling only the largest nodule in case of multinodular disease could lead to the risk of missing out significant features and non-matching with histological analysis: however, we opted for this approach in order to replicate the environment of a percutaneous biopsy, where the information is derived by a single more representative nodule.

Finally, considering the preliminary nature of this study, the models were not calibrated, and the discrimination ability of each one was not yet assessed, preferring a standard cut-off which could lead to an over or under estimation of the models' performance. However, the decision about how to set-up the cut-off should be further considered in the next steps of this work, depending on its clinical implications and not only on mathematical decisions.

## 5. Conclusions

In conclusion, our study introduces a paradigm shift in HCC research by presenting a highly effective MVI predictive model based solely on radiomics features. The departure from traditional clinical variables may represent a turning point in the automatization of preoperative staging. Future endeavors should aim to demystify the black box nature of machine learning models, ensuring seamless translation into clinical decision-making. Our results pave the way for a new era in liver cancer research, offering a valuable tool for clinicians to refine therapeutic strategies in the pursuit of enhanced patient outcomes.

## CRediT authorship contribution statement

**Simone Famularo:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft. **Camilla Penzo:** Formal analysis, Investigation, Methodology, Writing – original draft. **Cesare Maino:** Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Flavio Milana:** Data curation, Writing – original draft. **Riccardo Oliva:** Data curation, Writing – original draft. **Jacques Marescaux:** Funding acquisition, Writing – review & editing. **Michele Diana:** Investigation, Writing – original draft. **Fabrizio Romano:** Supervision, Writing – review & editing. **Felice Giuliante:** Supervision, Writing – review & editing. **Francesco Ardito:** Investigation, Writing – original draft. **Gian Luca Grazi:** Supervision, Writing – review & editing. **Matteo Donadon:** Conceptualization, Investigation. **Guido Torzilli:** Funding acquisition, Supervision, Writing – review & editing.

## Declaration of competing interest

No conflict of interest has to be reported for the present manuscript from each of the authors.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejso.2024.108274.

## References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clin 2021;71:209–49.
[2] Kanwal F, Befeler A, Chari RS, Marrero J, Kahn J, Afdhal N, et al. Potentially curative treatment in patients with hepatocellular cancer–results from the liver cancer research network. Aliment Pharmacol Ther 2012;36:257–65.
[3] Torzilli G, Belghiti J, Kokudo N, Takayama T, Capussotti L, Nuzzo G, et al. Reply to letter: "A snapshot of the effective indications and results of surgery for hepatocellular carcinoma in tertiary referral centers: is it adherent to the EASL/AASLD recommendations? An observational study of the HCC East-West study group": when the study setting "Ignores" the patients. Ann Surg 2015;262:e30–1.
[4] Famularo S, Donadon M, Cipriani F, Ardito F, Carissimi F, Perri P, et al. Hepatocellular carcinoma surgical and oncological trends in a national multicentric population: the HERCOLES experience. Updates Surg 2020. https://doi.org/10.1007/s13304-020-00733-6.
[5] Sapisochin G, Goldaracena N, Astete S, Laurence JM, Davidson D, Rafael E, et al. Benefit of treating hepatocellular carcinoma recurrence after liver transplantation and analysis of prognostic factors for survival in a large Euro-American Series. Ann Surg Oncol 2015;22:2286–94.
[6] Filgueira NA. Hepatocellular carcinoma recurrence after liver transplantation: risk factors, screening and clinical presentation. World J Hepatol 2019;11:261–72.
[7] Bodzin AS, Lunsford KE, Markovic D, Harlander-Locke MP, Busuttil RW, Agopian VG. Predicting mortality in patients developing recurrent hepatocellular carcinoma after liver transplantation: impact of treatment modality and recurrence characteristics. Ann Surg 2017;266:118–25.
[8] Erridge S, Pucher PH, Markar SR, Malietzis G, Athanasiou T, Darzi A, et al. Meta-analysis of determinants of survival following treatment of recurrent hepatocellular carcinoma. Br J Surg 2017;104:1433–42.
[9] Mazzaferro V, Llovet JM, Miceli R, Bhoori S, Schiavo M, Mariani L, et al. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: a retrospective, exploratory analysis. Lancet Oncol 2009;10:35–43.
[10] Lim K-C, Chow PK-H, Allen JC, Chia G-S, Lim M, Cheow P-C, et al. Microvascular invasion is a better predictor of tumor recurrence and overall survival following surgical resection for hepatocellular carcinoma compared to the Milan criteria. Ann Surg 2011;254:108–13.
[11] Sheng X, Ji Y, Ren G-P, Lu C-L, Yun J-P, Chen L-H, et al. A standardized pathological proposal for evaluating microvascular invasion of hepatocellular carcinoma: a multicenter study by LCPGC. Hepatol Int 2020;14:1034–47.
[12] Miyata R, Tanimoto A, Wakabayashi G, Shimazu M, Nakatsuka S, Mukai M, et al. Accuracy of preoperative prediction of microinvasion of portal vein in hepatocellular carcinoma using superparamagnetic iron oxide-enhanced magnetic resonance imaging and computed tomography during hepatic angiography. J Gastroenterol 2006;41:987–95.
[13] Rodríguez-Perálvarez M, Luong TV, Andreana L, Meyer T, Dhillon AP, Burroughs AK. A systematic review of microvascular invasion in hepatocellular carcinoma: diagnostic and prognostic variability. Ann Surg Oncol 2013;20:325–39.
[14] Hirokawa F, Hayashi M, Miyamoto Y, Asakuma M, Shimizu T, Komeda K, et al. Outcomes and predictors of microvascular invasion of solitary hepatocellular carcinoma. Hepatol Res 2014;44:846–53.
[15] Pawlik TM, Gleisner AL, Anders RA, Assumpcao L, Maley W, Choti MA. Preoperative assessment of hepatocellular carcinoma tumor grade using needle biopsy: implications for transplant eligibility. Ann Surg 2007;245:435–42.
[16] Lei Z, Li J, Wu D, Xia Y, Wang Q, Si A, et al. Nomogram for preoperative estimation of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma within the Milan criteria. JAMA Surg 2016;151:356–63.
[17] Renzulli M, Brocchi S, Cucchetti A, Mazzotti F, Mosconi C, Sportoletti C, et al. Can current preoperative imaging Be used to detect microvascular invasion of hepatocellular carcinoma? Radiology 2016;279:432–42.

[18] Xu X, Zhang H-L, Liu Q-P, Sun S-W, Zhang J, Zhu F-P, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. J Hepatol 2019;70:1133–44.
[19] Applying artificial intelligence to big data in hepatopancreatic and biliary surgery: a scoping review. Art Int Surg 2023;3:27–47.
[20] Singal AG, Mukherjee A, Elmunzer BJ, Higgins PDR, Lok AS, Zhu J, et al. Machine learning algorithms outperform conventional regression models in predicting development of hepatocellular carcinoma. Am J Gastroenterol 2013;108:1723–30.
[21] Kawka M, Dawidziuk A, Jiao LR, Gall TMH. Artificial intelligence in the detection, characterisation and prediction of hepatocellular carcinoma: a narrative review. Transl Gastroenterol Hepatol 2022;7:41.
[22] Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[23] Peng J, Zhang J, Zhang Q, Xu Y, Zhou J, Liu L. A radiomics nomogram for preoperative prediction of microvascular invasion risk in hepatitis B virus-related hepatocellular carcinoma. Diagn Interv Radiol 2018;24:121–7.
[24] World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310:2191–4.
[25] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med 2007;4:e296.
[26] Nioche C, Orlhac DF, Boughdad S, Reuzé S, Goya-Outi J, Robert C, Pellot-Barakat C, Soussan M, Frouin F, Buvat I. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. Cancer Res 2018;78(16):4786–9.
[27] van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained Equations inR. J Stat Software 2011;45. https://doi.org/10.18637/jss.v045.i03.
[28] White IR, Royston P. Imputing missing covariate values for the Cox model. Stat Med 2009;28:1982–98.
[29] Famularo S, Piardi T, Molfino S, Di Martino M, Ferrari C, Ielpo B, et al. Factors affecting local and intra hepatic distant recurrence after surgery for Hcc: an alternative perspective on microvascular invasion and Satellitosis - a Western European multicentre study. J Gastrointest Surg 2021;25:104–11.
[30] Famularo S, Ceresoli M, Giani A, Ciulli C, Pinotti E, Romano F, et al. Is it just a matter of surgical extension to achieve the cure of hepatocarcinoma? A meta-analysis of propensity-matched and randomized studies for anatomic versus parenchyma-sparing liver resection. J Gastrointest Surg 2021;25:94–103.
[31] Qin S, Chen M, Cheng A-L, Kaseb AO, Kudo M, Lee HC, et al. Atezolizumab plus bevacizumab versus active surveillance in patients with resected or ablated high-risk hepatocellular carcinoma (IMbrave050): a randomised, open-label, multicentre, phase 3 trial. Lancet 2023;402:1835–47.
[32] Xia T-Y, Zhou Z-H, Meng X-P, Zha J-H, Yu Q, Wang W-L, et al. Predicting microvascular invasion in hepatocellular carcinoma using CT-based radiomics model. Radiology 2023;307:e222729.
[33] Ma X, Wei J, Gu D, Zhu Y, Feng B, Liang M, et al. Preoperative radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using contrast-enhanced CT. Eur Radiol 2019;29:3595–605.
[34] Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol 2007;25:675–80.
[35] Feng L-H, Dong H, Lau W-Y, Yu H, Zhu Y-Y, Zhao Y, et al. Novel microvascular invasion-based prognostic nomograms to predict survival outcomes in patients after R0 resection for hepatocellular carcinoma. J Cancer Res Clin Oncol 2017;143:293–303.
[36] Yusa T, Yamashita Y-I, Okabe H, Nakao Y, Itoyama R, Kitano Y, et al. Survival impact of immune cells infiltrating peritumoral area of hepatocellular carcinoma. Cancer Sci 2022;113:4048–58.
[37] Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inf Decis Making 2020;20:1–9.