



OPEN

## Contrast phase recognition in liver computer tomography using deep learning

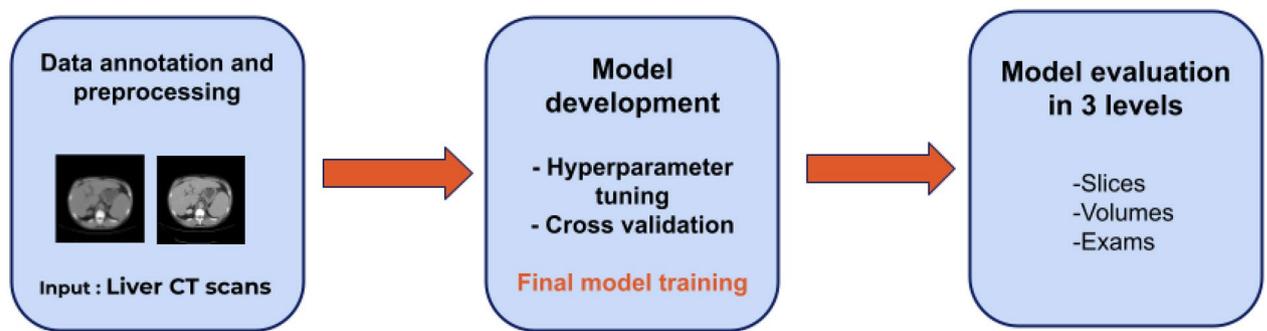
Bruno Aragão Rocha<sup>1,2</sup>✉, Lorena Carneiro Ferreira<sup>1</sup>, Luis Gustavo Rocha Vianna<sup>2,5</sup>, Luma Gallacio Gomes Ferreira<sup>2,5</sup>, Ana Claudia Martins Ciconelle<sup>2</sup>, Alex Da Silva Noronha<sup>2</sup>, João Martins Cortez Filho<sup>4</sup>, Lucas Salume Lima Nogueira<sup>4</sup>, Jean Michel Rocha Sampaio Leite<sup>2</sup>, Maurício Ricardo Moreira da Silva Filho<sup>1</sup>, Cláudia da Costa Leite<sup>1</sup>, Marcelo de Maria Felix<sup>1</sup>, Marco Antônio Gutierrez<sup>3</sup>, Cesar Higa Nomura<sup>1</sup>, Giovanni Guido Cerri<sup>1</sup>, Flair José Carrilho<sup>4</sup> & Suzane Kioko Ono<sup>4</sup>

Hepatocellular carcinoma (HCC) has become the 4th leading cause of cancer-related deaths, with high social, economical and health implications. Imaging techniques such as multiphase computed tomography (CT) have been successfully used for diagnosis of liver tumors such as HCC in a feasible and accurate way and its interpretation relies mainly on comparing the appearance of the lesions in the different contrast phases of the exam. Recently, some researchers have been dedicated to the development of tools based on machine learning (ML) algorithms, especially by deep learning techniques, to improve the diagnosis of liver lesions in imaging exams. However, the lack of standardization in the naming of the CT contrast phases in the DICOM metadata is a problem for real-life deployment of machine learning tools. Therefore, it is important to correctly identify the exam phase based only on the image and not on the exam metadata, which is unreliable. Motivated by this problem, we successfully created an annotation platform and implemented a convolutional neural network (CNN) to automatically identify the CT scan phases in the HCFMUSP database in the city of São Paulo, Brazil. We improved this algorithm with hyperparameter tuning and evaluated it with cross validation methods. Comparing its predictions with the radiologists annotation, it achieved an accuracy of 94.6%, 98% and 100% in the testing dataset for the slice, volume and exam evaluation, respectively.

Over the past few years, diagnostic medicine has achieved prominent success in the development of tools based on machine learning (ML) algorithms, especially by deep learning techniques such as convolutional neural networks (CNN), which are the most used and suitable approach for imaging analysis<sup>1</sup>. This computer-oriented approach enables the extraction of imaging patterns based on the ability to learn from data and has been successfully applied in diagnosis and treatment of several conditions, including melanoma, nail mycosis, pneumonia, acute respiratory distress syndrome (ARDS) and coronavirus disease, as well as in liver diseases, such as hepatocellular carcinoma (HCC)<sup>2-7</sup>.

Hepatocellular carcinoma has become the 4th leading cause of cancer-related deaths, with an increasing incidence, especially in western nations<sup>8-10</sup>. Magnetic resonance imaging (MRI) and multiphase computed tomography (CT) are currently the gold standard imaging method for detecting HCC with no need for a biopsy if a typical imaging pattern is present i.e. whenever a mass measuring 1 cm is found that demonstrates arterial hyper-enhancement and one or more major features in selected patients with high risk for HCC, according to the Liver Imaging Reporting and Data System<sup>8,11-14</sup>. These techniques involve intravenous contrast injection with a four-phase image acquisition protocol (unenhanced, arterial, portal and delayed), which is the clinical reference standard.

<sup>1</sup>InRad, Institute of Radiology, University of São Paulo, School of Medicine, Rua Dr. Ovídio Pires de Campos, 75 Cerqueira César, São Paulo, SP 05403-010, Brazil. <sup>2</sup>Machiron Ltd., Rua Capote Valente, 671, São Paulo 05409-002, Brazil. <sup>3</sup>Informatics Department, The Heart Institute, Hospital das Clínicas (HCFMUSP), University of São Paulo, School of Medicine, Rua Dr. Enéas de Carvalho Aguiar 44, São Paulo, SP 05403-000, Brazil. <sup>4</sup>Department of Gastroenterology, University of São Paulo, School of Medicine (FMUSP), Hospital das Clínicas (HCFMUSP), Rua Dr. Enéas Carvalho de Aguiar, 225, São Paulo, SP 05403-000, Brazil. <sup>5</sup>These authors contributed equally: Luis Gustavo Rocha Vianna and Luma Gallacio Gomes Ferreira. ✉email: bruno@machiron.com.br



**Figure 1.** Overview of the methodology and application. Our methodology is divided into three main steps: (1) data annotation and preprocessing, (2) model development and (3) model evaluation.

Nonetheless, liver segmentation and HCC identification pipelines present many challenges related to the input images quality. For instance, Derclé et al.<sup>15</sup> demonstrated that the quality of CT scans acquired at the portal venous phase was suboptimal in one third of colorectal cancer patients. In addition, although radiologists can easily recognize the study phase visually, automatic phase identification is important for the deployment of future HCC screening algorithms. In the best scenario, automatic identification of CT scan phases should not be a problem, since there is a field called SeriesDescription in the DICOM metadata<sup>16</sup>, which has the name of the acquired series. However, in a real-life and clinical setting, the main problem is that there is no standard for naming across machines, even on machines of the same manufacturer. The series acquisition time DICOM Tag could also be a way to identify the series phase. However, this approach would only work if all exams had all four phases in clinical practices. As many exams may have fewer phases (for example, only the unenhanced and portal phase), this reinforces the importance of identifying the phase directly by analyzing the image.

The lack of standardization and proper quality control of contrast-enhanced phases in abdominal CT scans is a pivotal limitation in the field of radiomics<sup>17</sup>, and, to the best of our knowledge, very few attempts have been made to address these issues<sup>18</sup>.

Hence, these problems motivated this group to implement an algorithm to automatically classify the CT scans phases in Hospital das Clínicas (HCFMUSP) database in São Paulo, Brazil. We created the Liver Artificial Intelligence (HepatiA) platform to store the CT scans and facilitate the radiologists annotation of the exams. Using an unprocessed DICOM folder containing all four contrast phases as input, we implemented a fully-automated CNN algorithm that outputs a sorted folder for each of the relevant contrast phases.

**Organization of article.** The introduction of this article describes our motivation to identify the contrast phases in a CT scan. In the following section, there is an overview of our methodology and a description of HepatiA platform. In “Methods” section, we present a detailed description of data collected, annotation process, algorithm construction and evaluation methods. Then, we present the results, discussion and conclusion. After the references, we added the acknowledgements, author contributions statement, additional information about the dataset and competing interests.

## Overview of the methodology

As shown in Fig. 1, the data used in this paper are abdominal CT scans acquired with a protocol for evaluation of the liver, which can be evaluated in three levels: (1) slice, which is a single image of the exam; (2) volume, which is the group of slices that belongs to a single contrast phase; and (3) exam, which includes the four volumes from the corresponding four phases. The training step was performed in two steps. Firstly, we performed a hyperparameter tuning to find the best combination of parameters for the convolution neural network (CNN). Secondly, we trained the defined model with our dataset. For a better evaluation of the final structure of our model, we included a cross validation step. The developed CNN analyzes slices, generating the probability of the slice belonging to each phase. For the analysis of volumes and exams, we combined the results from the slices with a post-processing technique. More details of the dataset and methods used is described in “Methods” section.

## Methods

**Study population.** The study was approved by Ethics Committee at the *Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo* under the study protocol CAAE 69385217.1.0000.0068 in accordance with the ethical guidelines of the 1975 Declaration of Helsinki. The need for written informed consent was formally waived by Ethics Committee at the *Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo* due to its retrospective, single-center nature. Our study population is composed of 396 CT scans from unique patients, which are composed with 4 volumes each, summing in 178.633 slices. The data are from healthy liver donors (20%) and cirrhotic patients (80%), from the Division of Clinical Gastroenterology and Hepatology of the Hospital das Clínicas at the University of São Paulo (HCFMUSP) in São Paulo, SP, Brazil, from 2008 to 2021. These patients underwent abdominal Multi-phase Contrast-enhanced Computed Tomography (CT) to assess liver conditions using a 4-phase protocol (unenhanced, arterial, portal-venous and delayed).



**Figure 2.** Demonstration of the contrast phases of a liver CT scan.

**Figure 3.** HepatIA platform liver tomography contrast phase annotation screen.

**Database and preprocessing.** As our exams are obtained with a four-phase image acquisition protocol, the exam's volumes should be labeled indicating the phase, which are: unenhanced phase (also known as non-enhanced) is an image acquisition before the administration of intravenous contrast; Arterial phase is an acquisition about 35–45 s after intravenous contrast injection; portal phase is 60–75 s post-injection and delayed phase is about 3 minutes post-injection (Fig. 2).

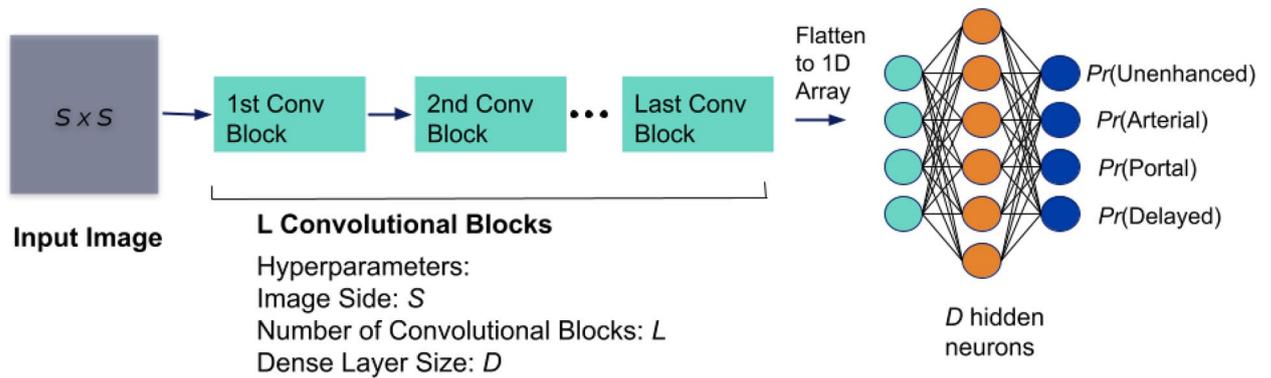
For the creation of the database and dataset preparation, we created a web-based platform called HepatIA implemented with Django 3.2.9, where the CT scans were stored in an Orthanc DICOM server (version 1.5.8) which is connected to the hospital PACS, while patient and clinical-related information was stored in a PostgreSQL relational database<sup>19</sup>.

Our platform allows the radiologists to access and fill out exam information, including the correct labels for each of the four contrast phases. This phase annotation was performed by three radiologists with two, four and eleven years of experience, by subjective analysis of the images using the DICOM viewer tool integrated into the HepatIA web-based platform (Fig. 3). Each reader checked the written phase DICOM tag and looked directly at the image to confirm it. We considered this activity to have a very low probability of human error, therefore, we dispensed the need for double reading.

A total of 396 exams (CT scans), from unique patients, were collected in DICOM format. The scans were randomly split into training (80%) and testing (20%). For each CT scan, we used all the four phases and, for each phase, we selected up to 150 slices that were spaced approximately evenly. The exams were taken by different CT machines, of which 214, 91, 90 and 2 are from Philips, GE Medical Systems, Siemens Health and Toshiba, respectively.

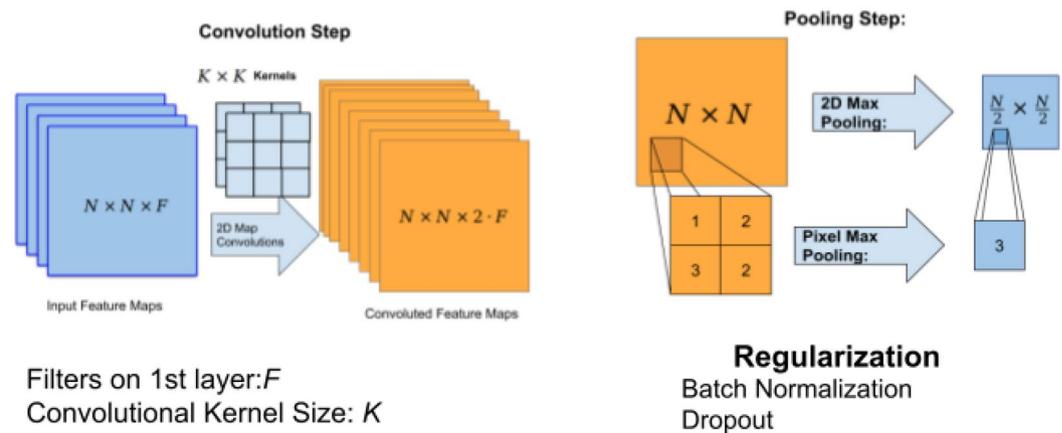
**Proposed model.** The implemented model was a Convolutional Neural Network (CNN)<sup>20</sup> with a dense final classification layer. Figure 4 shows the high level description of the architecture which consists of a series of convolutional blocks with an additional densely connected hidden layer and one output for each of the four contrast phases.

## Convolutional Neural Network (CNN)



**Figure 4.** High level description of the CNN architecture.

## Convolutional Block



**Figure 5.** Convolutional block with the convolutional and pooling steps, followed by regularization.  $N$  input feature size.

The convolutional blocks that we used are detailed on Fig. 5. Each block consists of a convolutional step, which aims to identify patterns in the images, and a pooling step for reducing the image dimensions, and, finally, a regularization step, where batch normalization and dropout are applied.

**Hyperparameter tuning.** The definition of the model depends on several parameters, such as number of convolutional layers, and finding the best combination is not a trivial task. As a solution, we used the Hyperband parameter tuning algorithm<sup>21</sup> to test the hyperparameters described in Table 1 for all possible values indicated on the table.

The Hyperband algorithm was chosen since it can test a wide range of hyperparameter sets with few epochs and only does full time training on the most promising combinations, thus allowing many more values to be tested within a reasonable tuning time.

In the case of Dropout and Learning Rate, the tuning algorithm samples values according to a log-based probability distribution, assigning equal probabilities to each order of the magnitude range. The Learning Epochs and Batch values were set after testing for impacts on execution time and memory consumption.

For this step, we used the train data set split into training (80%) and validation (20%). The evaluation of the model with the chosen hyperparameters was made with the test data set. The final parameters of the model are presented in the Results section.

**Cross-validation.** Using the data set that was previously assigned for training purposes and the model defined with the tuning procedure, we performed a  $k$ -fold cross-validation, where  $k$  corresponds to both the number of subsets the data was randomly sorted into, as well as the number of iterations. In this investigation, we set  $k$  as 5. For each iteration, four of the folds were used for training the CNN while the remaining one was used for testing, so that, at the end, each input data (either slices, volumes or exams) was used exactly four times for training and once for testing, respectively.

Hyperparameter		
Class	Name	Possible values
Architecture	Convolutional layers	2, 3, 4, 5, 6
Architecture	Convolutional kernel	$3 \times 3, 5 \times 5, 7 \times 7$
Architecture	Filters on 1st layer	8, 16
Architecture	Dense layer size	8, 16, 24, 32
Architecture	Image side	256, 512
Regularizer	Dropout	[1e-6, 1e-2]
Regularizer	Augmentation	Yes or no
Regularizer	Batch normalization	Yes or no
Optimizer	Optimizer	Adam, SGD
Optimizer	Learning rate	[1e-6, 1e-2]

**Table 1.** Hyperparameters of the CNN.

**Loss function.** In order to measure and minimize the errors over the epochs during the training and validation steps, we used the Categorical Cross-Entropy loss function, defined as:

$$\text{Categorical\_Cross\_Entropy}(o, y, p) = - \sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad (1)$$

where  $y_{o,c}$  corresponds to a binary indicator, i.e. 1 if class label  $c$  is the correct classification for this observation and 0 otherwise.  $M$  corresponds to the classes, which are the four exam phases; Observation  $o$  is the slice;  $p_{o,c}$  is the model predicted probability that observation  $o$  is from class  $c$ .

**Evaluation. Prediction levels.** The algorithm generates predictions in three evaluation levels: individual slices, individual phase volumes and exams. The model analyzes the 3D input images by processing one 2D slice at a time. The algorithm assigns a score to each input slice for all possible phases. Hence, volume-level prediction is made by choosing the corresponding phase with the highest mean score across all slices. Finally, the full exam prediction is made by combining the predictions of its corresponding volumes. Since there is exactly one volume for each phase, in cases where more than one volume has the same volume-level prediction the volume with highest confidence is chosen and the other volume's prediction is changed to the next available option.

For both slice and volume levels several classification performance metrics are described in the next section. Exam level predictions were evaluated according to the number of volumes correctly identified.

**Evaluation metrics.** Herein, we evaluated the performance of our model in the testing set using the most common metrics for classification problems: F1 score, Area under the ROC curve, accuracy, precision and recall<sup>22</sup>. These metrics were calculated in a one-vs-rest manner .i.e. there will be four measures for each evaluation metric, each of them corresponding to correctly classifying the images as a given class against all the other classes. The means were also reported for each evaluation metric.

The accuracy of the model is calculated as the fraction of images classified correctly over the total number of images:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

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

The following metrics are defined based on a class  $C$ , in this case, the classes are the four phases of the CT scan. The precision is defined as follows:

$$\text{Precision}(C) = \frac{TP(C)}{TP(C) + FP(C)} \quad (3)$$

while the recall (also referred to sensitivity or true positive rate) corresponds to:

$$\text{Recall}(C) = \frac{TP(C)}{TP(C) + FN(C)} \quad (4)$$

The formula for the F1 score is defined as:

$$\text{F1score}(C) = \frac{2 \times \text{Precision}(C) \times \text{Recall}(C)}{\text{Precision}(C) + \text{Recall}(C)} \quad (5)$$

As the last metric, AUC is the area under the ROC curve, which is the plot of FPR x recall. FPR is given by:

Hyperparameter		
Class	Name	Value
Architecture	Convolutional layers	5
Architecture	Convolutional kernel	5 × 5
Architecture	Filters on 1st layer	32
Architecture	Dense layer size	32
Architecture	Image side	256
Regularizer	L2 *	0.01
Regularizer	Dropout	2.045 × 10 <sup>-4</sup>
Regularizer	Augmentation	No
Regularizer	Batch normalization	Yes
Optimizer	Optimizer	Adam
Optimizer	Learning rate	1.483 10 <sup>-3</sup>
Optimizer	Loss function *	Cross entropy
Execution	Learning epochs *	50
Execution	Batch *	5

**Table 2.** Best hyperparameters after tuning. \*Only one value was tested for this hyperparameter.

$$\text{False\_positive\_rate}(C) = \frac{FP(C)}{FP(C) + TN(C)} \quad (6)$$

Each of these metrics is calculated per slice prediction as well as per volume prediction.

Lastly, the performance of the analysis per exam was expressed in terms of how many phases in each exam were correctly classified. In this step, we iteratively associate volumes with phases. The first assignment of a volume-phase pair is done according to the highest prediction among all possible volume-phase pairings. After that, the next highest prediction value for a pairing that does not contain the already paired volume and phase is identified and the corresponding volume and phase are paired. This step will happen until only one volume and one phase remain to be paired. Thus, in the end, an exam will always have each of its 4 volumes associated with a different phase as well as an indicated volume for each of the four phases.

The implementation was performed using Python 3.8 environment with Tensorflow 2.6 and Keras 2.6 packages. The computer used had the following specifications: Linux Ubuntu 20.04 Virtual Machine in a FOXCONN M100-NHI High Processor Computing (HPC) with 32 Cores CPU @ 2.9 GHz 346 GB RAM and a cluster of 16 NVIDIA Tesla V100 16 GB cards.

## Results

**Hyperparameter tuning.** Table 2 shows best values for the hyperparameters adjusted with hyperparameter tuning. These values were then used in all further analysis.

Comparing a model trained with the default hyperparameter values with a model trained with the best tuned values, we observed an accuracy increase from 78.02 to 85.49% in the slices evaluation and 89.90 to 92.31% in the volume one.

**Best model results.** Figure 6 displays the accuracy and the loss of the model for both training and validation data sets over 25 epochs, that is, 25 rounds of evaluation with all training images. The highest accuracy with the smallest loss was achieved in the 23rd epoch in the validation data set.

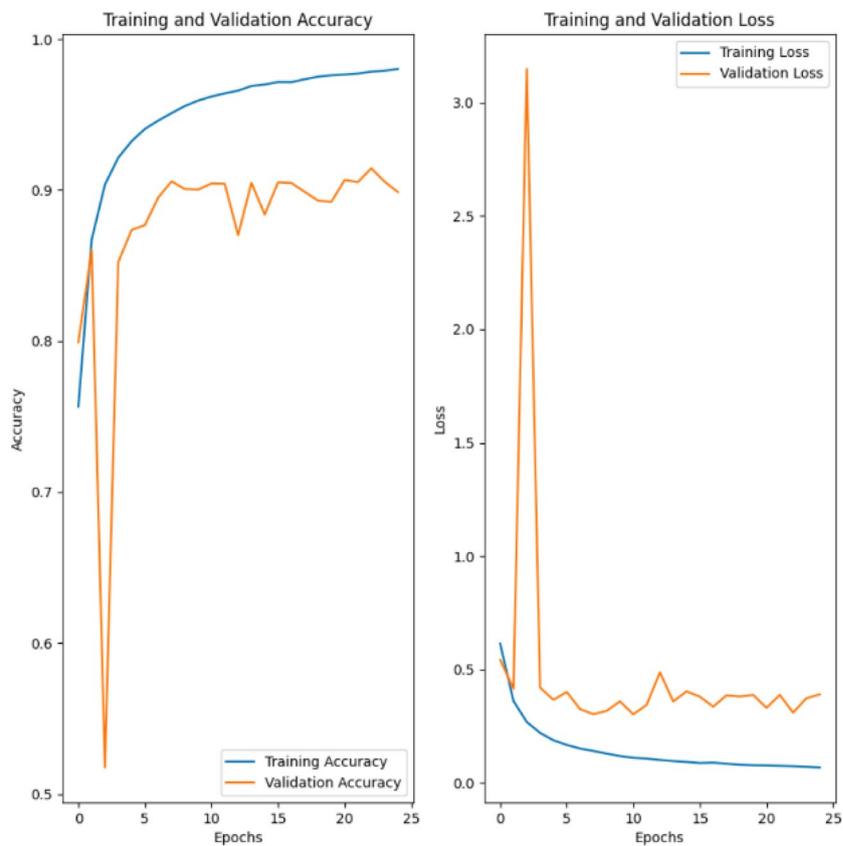
The output of the model is the probabilities of each slice belonging to each contrast phase, as shown in Fig. 7. The final classification of the slice is the contrast phase with higher probability. The accuracy achieved in the testing set was 94.6%. Among the other metrics, as shown in Table 3, the unenhanced phase had the best results.

When combining each slice prediction into a single volume prediction, the accuracy rose to 98%. Increases were also observed for the other evaluation metrics, as shown in Table 4. Across the four phases, the F1-score and AUC metrics show a decrease in performance according to the order of occurrence of the phases. Particularly, the unenhanced has the best values while the delayed has the lowest ones. It is possible to observe in Fig. 8 that most mistakes in classification are in the portal and delayed phases.

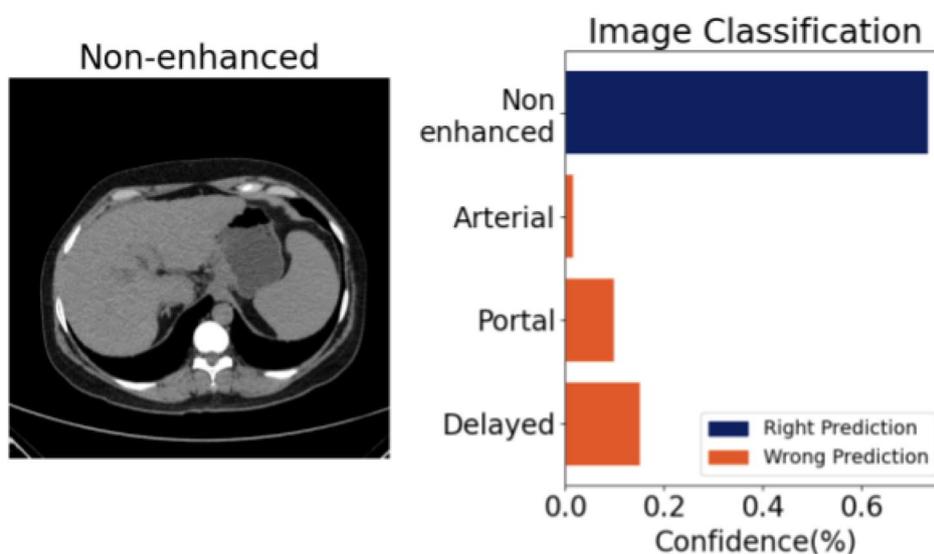
The accuracy of the exams evaluation was 100%, which means that all phases in all exams were correctly classified.

**Cross validation results.** For a better evaluation of our model, we added a cross validation step using four folds. Despite the decreasing performance shown in Tables 5 and 6, the general accuracy for slices and volumes was above 92%. The accuracy for the exams was 96.5 ± 3.93%.

Overall, all metrics were higher in the volumes evaluation than in the slices one. For both slices and volumes, the highest metrics were achieved in the unenhanced and arterial phases.



**Figure 6.** Accuracy and loss in the training and validation data sets of a CNN to identify the four phases of abdominal CT scans.



**Figure 7.** An example of slice with correct prediction of contrast phase. This plot shows the confidence of the slice belonging to each contrast phase. For this slice, the correct phase is *non-enhanced* and the algorithm classified it correctly with 70% of confidence, while it also gave a probability of approximately 18% of belonging to delayed, 10% of belonging to portal and 2% of belonging to arterial.

	Precision	Recall	F1-score	AUC	N
Unenhanced	0.99	0.99	0.99	1	10,757
Arterial	0.98	0.98	0.98	1	10,870
Portal	0.92	0.89	0.89	0.99	11,907
Delayed	0.89	0.93	0.91	0.99	10,803
Average	0.95	0.95	0.95	0.99	–

**Table 3.** Performance metrics in the testing set for the slices evaluation using the best model found by the Hyperparameter tuning.

	Precision	Recall	F1-score	AUC	N
Unenhanced	1.00	1.00	1.00	1.00	80
Arterial	0.99	1.00	0.99	1.00	80
Portal	0.97	0.99	0.98	1.00	80
Delayed	0.99	0.96	0.97	0.99	80
Average	0.99	0.99	0.99	0.99	–

**Table 4.** Performance metrics in the testing set for the volumes evaluation using the best model found by the Hyperparameter Tuning.

	Precision	Recall	F1-score	AUC	N
Unenhanced	0.98 ± 0.02	0.99 ± 0.01	0.99 ± 0.01	1 ± 0.00	10,839.25 ± 126.79
Arterial	0.97 ± 0.01	0.95 ± 0.04	0.96 ± 0.02	0.99 ± 0.01	10,972.75 ± 148.56
Portal	0.87 ± 0.06	0.87 ± 0.04	0.87 ± 0.02	0.98 ± 0.01	11,770.0 ± 129.03
Delayed	0.89 ± 0.03	0.88 ± 0.05	0.88 ± 0.01	0.98 ± 0.00	10,926.25 ± 174.68
Average	0.93 ± 0.01	0.92 ± 0.01	0.93 ± 0.01	0.99 ± 0.00	–

**Table 5.** Performance metrics in the testing set for the slices evaluation using cross validation. The values are mean ± standard deviation.

	Precision	Recall	F1-score	AUC	N
Unenhanced	0.99 ± 0.01	1.00 ± 0.00	0.99 ± 0.01	1.00 ± 0.00	80
Arterial	0.99 ± 0.01	0.97 ± 0.04	0.98 ± 0.02	0.99 ± 0.01	80
Portal	0.93 ± 0.05	0.97 ± 0.02	0.95 ± 0.03	0.99 ± 0.01	80
Delayed	0.97 ± 0.02	0.94 ± 0.02	0.96 ± 0.02	0.99 ± 0.00	80
Average	0.97 ± 0.02	0.97 ± 0.02	0.97 ± 0.02	0.99 ± 0.00	–

**Table 6.** Performance metrics in the testing set for the volumes evaluation using cross validation. The values are mean ± standard deviation.

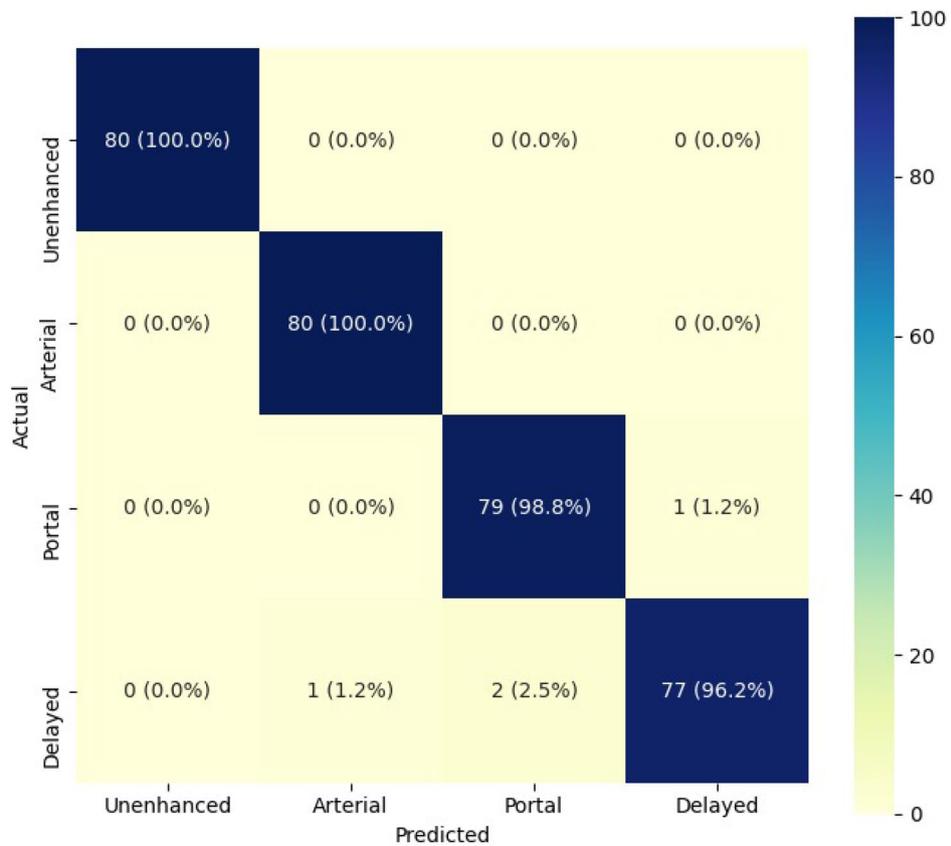
Given that the exams used in this experiment are from four different CT machine manufacturers, we evaluated the accuracy of the exams for each manufacturer. The results show that there was not a significant difference among them and the accuracy is above 95% for all machines.

In addition, exams from patients with chronic liver disease ( $n = 65$ ) have an accuracy of  $96.5 \pm 3.4\%$ , while the exams from healthy patients ( $n = 15$ ) have  $96.7 \pm 6.7\%$ .

## Discussion

Herein, the CNN model developed to predict, identify and differentiate contrast-enhanced phases in liver computer tomography presented results with high accuracy.

Besides the accuracy achieved using each single slice as input being slightly higher than the one achieved by Derclé et al.<sup>18</sup>, who used a random forest classifier for predicting optimal-portal phase (85% in our analysis compared to 84% in theirs), the average accuracy of our algorithm rose to 98% in the volumes evaluation. In addition, our algorithm's performance rose up to 99% when considering other evaluation metrics such as the



**Figure 8.** Confusion matrix heat map of the testing set volumes.

AUC ( $0.997 \pm 0.002$ ). Hence, this result further demonstrates the potential of our approach to be applied in the clinical setting. In terms of the delayed phase's accuracy, our findings are also in unison with Derclé et al.<sup>18</sup>, since this phase concentrated most of prediction mistakes among all the evaluated phases.

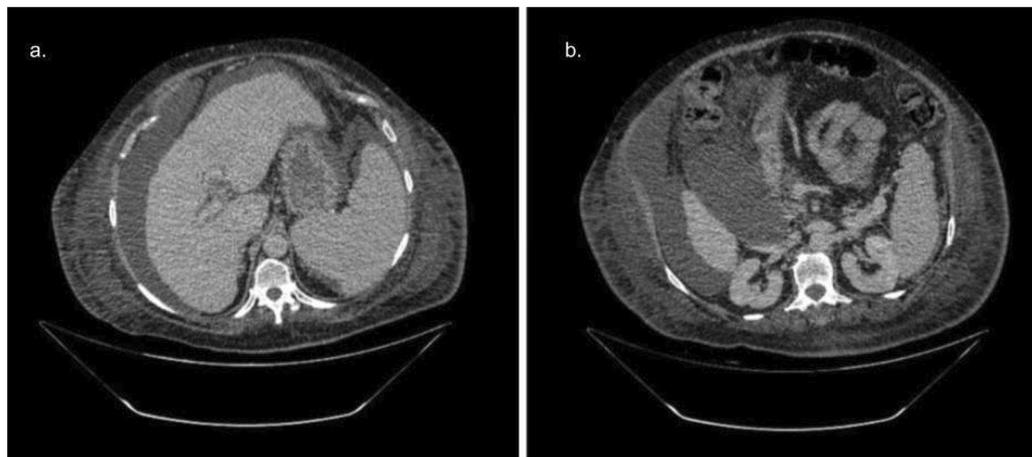
Nonetheless, it should be noted that these findings may be compared with some caution, since the input, algorithm and output data differ among our study and the one by Derclé et al.<sup>18</sup>. For instance, we aimed to precisely identify only four phases of the exams (unenhanced, arterial, portal and delayed), since the CT evaluation of the liver is most commonly based on this four-phase protocol rather than a five-phase one that includes an optimal-portal category described by Derclé et al.<sup>18</sup>. Furthermore, we developed a fully-automated solution using only imaging data as input for our CNN algorithm, as opposed to Derclé et al.<sup>18</sup>, who used the mean intensities of the abdominal aorta and portal vein extracted from specialist annotated pixels.

The success of our approach is also evidenced by the performance of the exams evaluation, which, to the best of our knowledge, had not been previously tested by any research group. Furthermore, we included a hyperparameter tuning step, which has been shown to slightly improve the accuracy of AI algorithms, being fundamental for any cutting-edge pipeline<sup>23</sup>. In addition, we showed that there is no variability among CT machines in terms of algorithms' accuracy. Furthermore, one of the advantages of our approach does not need to use metadata in order to be applied to real data sets.

In the test set volume, the algorithm misclassified the exam series in four volumes. By analyzing each case individually, the reason behind the misclassification cannot be determined with certainty, but it is possible to infer potential confounding factors. In two cases, the algorithm classified a delayed phase as a portal as shown in Fig. 9. When analyzing the images, we observed that these correspond to patients whose delayed phase showed no contrast excreted in the collecting system. We deduced that, probably, the lack of contrast in the collecting system in these patients with possible slow cardiac output or slow renal excretion rate may be the confounding factor. In this kind of situation, a radiologist who only has access to this phase of the exam will also have difficulties in differentiating whether it is a delayed phase or a portal phase with poor quality. In the other two cases, the algorithm classified a portal phase as an delayed phase as shown in Fig. 10. There was extensive thrombosis of the trunk of the portal vein, causing the liver and the portal vein not to present significant enhancement, which may have been the confounding factor.

An important aspect to be taken into account in this study is the fact that it included healthy patients and mostly cirrhotic patients, including in very advanced stages. This composition of the patient profile shows that the algorithm performs well even in a population with a high prevalence of liver disease.

For instance, in a clinical environment, this kind of deep learning approach to identify contrast phase may also be useful for creating databases that facilitate radiologists' daily routine and, ultimately, accelerate the diagnosis



**Figure 9.** Two slices from the delayed phase incorrectly classified as portal phase. In general, in the delayed phase, we notice contrast being excreted in the urinary collecting system. In this case, despite being a delayed phase, no contrast was observed, probably due to low cardiac output or slow urinary function.



**Figure 10.** Example of a portal phase incorrectly classified as a delayed phase. Note that there is an infiltrative hepatocellular carcinoma with extensive portal vein thrombosis.

of liver diseases, identifying, selecting and organizing images within the DICOM viewer, based on post-contrast phases, saving one more step in the analysis. Besides, it could also allow the post-contrast images' quality evaluation, bringing new possibilities for quality control tools.

Furthermore, a CT phase identifier can be used as a general, yet extremely useful strategy to organize and clean the input data for any algorithm that utilizes image information. As noted by Castaldo et al.<sup>1</sup>, these two issues are among the main limitations in the fields of radiomics and artificial intelligence. In this context, this AI solution can potentially aid in solving those issues and might be incorporated in other radiology-pipelines, improving the accuracy of virtually any CT abdominal algorithm that uses specific contrast phases as input.

These results show a promising application of a CNN-based phase identifier. However, it is necessary to evaluate it on a larger dataset of exams, especially with more diverse machines and image acquisition protocols to guarantee the maintenance of performance and subsequent practical implementation of this type of approach.

## Conclusion

Our study successfully demonstrates an approach to phase recognition of contrast-enhanced abdominal CT scans using a Convolutional Neural Network (CNN). Considering the high mean accuracy of our algorithm, our results can significantly aid in the proper and fundamental standardization and quality control of input data for liver segmentation and identification of HCC lesions. Further validation studies are necessary to check whether the algorithm's accuracy remains the same for a diverse set of CT exams, especially for different patient profiles and institutions with potentially different infusion pumps and contrast timing protocols. Addressing this question will be fundamental to tailor AI solutions for a faster and more proper care of patients with liver diseases.

## Data availability

The dataset generated during the current study is not publicly available due to institutional privacy policies but is available from the corresponding author on reasonable request.

Received: 13 June 2022; Accepted: 16 November 2022

Published online: 24 November 2022

## References

- Castaldo, A. *et al.* State of the art in artificial intelligence and radiomics in hepatocellular carcinoma. *Diagnostics* **11**, 1194. <https://doi.org/10.3390/diagnostics11071194> (2021).
- Ferreira Junior, J. R. *et al.* Novel chest radiographic biomarkers for COVID-19 using radiomic features associated with diagnostics and outcomes. *J. Digit. Imaging* **34**, 297–307. <https://doi.org/10.1007/s10278-021-00421-w> (2021).
- Graves, C. V., Moreno, R. A., Rebelo, M. S., Nomura, C. H. & Gutierrez, M. A. Improving the generalization of deep learning methods to segment the left ventricle in short axis MR images. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1203–1206. <https://doi.org/10.1109/EMBC44109.2020.9175256> (IEEE, 2020).
- Brinker, T. J. *et al.* A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *Eur. J. Cancer* **111**, 148–154. <https://doi.org/10.1016/j.ejca.2019.02.005> (2019).
- Han, S. S. *et al.* Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS One* **13**, 1–14. <https://doi.org/10.1371/journal.pone.0191493> (2018).
- Erickson, B. J., Korfiatis, P., Akkus, Z. & Kline, T. L. Machine learning for medical imaging. *Radiographics* **37**, 505–515. <https://doi.org/10.1148/rg.2017160130> (2017).
- Rajpurkar, P. *et al.* CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, pp 3–9. [arXiv:1711.05225](https://arxiv.org/abs/1711.05225) (2017).
- Yang, J. D. *et al.* A global view of hepatocellular carcinoma: Trends, risk, prevention and management. <https://doi.org/10.1038/s41575-019-0186-y> (2019).
- Waller, L. P., Deshpande, V. & Pyporopoulos, N. Hepatocellular carcinoma: A comprehensive review. *World J. Hepatol.* **7**, 2648–2663. <https://doi.org/10.4254/wjh.v7.i26.2648> (2015).
- Paranáguá-Vezozzo, D. C. *et al.* Epidemiology of HCC in Brazil: Incidence and risk factors in a ten-year cohort. *Ann. Hepatol.* **13**, 386–93 (2014).
- Sirlin, C. B. LI-RADS® Diagnostic Categories. In *CT/MRI Manual*, chap. 8 (2018).
- Marrero, J. A. *et al.* Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* **68**, 723–750. <https://doi.org/10.1002/hep.29913> (2018).
- Galle, P. R. *et al.* EASL clinical practice guidelines: Management of hepatocellular carcinoma. *J. Hepatol.* **69**, 182–236. <https://doi.org/10.1016/j.jhep.2018.03.019> (2018).
- Bruix, J. & Sherman, M. Management of hepatocellular carcinoma: An update. *Hepatology* **53**, 1020–1022. <https://doi.org/10.1002/hep.24199> (2011).
- Derclé, L. *et al.* Impact of variability in portal venous phase acquisition timing in tumor density measurement and treatment response assessment: Metastatic colorectal cancer as a paradigm. *JCO Clin. Cancer Inform.* <https://doi.org/10.1200/CCI.17.00108> (2017).
- Bidgood, W. D., Horii, S. C., Prior, F. W. & Van Syckle, D. E. Understanding and using DICOM, the data interchange standard for biomedical imaging. *J. Am. Med. Inform. Assoc.* **4**, 199–212. <https://doi.org/10.1136/jamia.1997.0040199> (1997).
- Miranda Magalhaes Santos, J. M. *et al.* State-of-the-art in radiomics of hepatocellular carcinoma: A review of basic principles, applications, and limitations. *Abdominal Radiol.* **45**, 342–353. <https://doi.org/10.1007/s00261-019-02299-3> (2020).
- Derclé, L. *et al.* Using a single abdominal computed tomography image to differentiate five contrast-enhancement phases: A machine-learning algorithm for radiomics-based precision medicine. *Eur. J. Radiol.* <https://doi.org/10.1016/j.ejrad.2020.108850> (2020).
- Ciconelle, A. C. M. *et al.* Database design and implementation of a convolutional neural network (CNN) for liver segmentation. *Hepatology* **72**(suppl.1), 677A (2020).
- Albawi, S., Mohammed, T. A. M. & Alzawi, S. Layers of a convolutional neural network. *Ieee* (2017).
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A. & Talwalkar, A. Hyperband: A novel bandit-based approach to hyperparameter optimization. *J. Mach. Learn. Res.* **18**, 1–52 (2018).
- Erickson, B. J. & Kitamura, F. Magician's corner: 9. performance metrics for machine learning models. *Radiol. Artif. Intell.* **3**, 1–7. <https://doi.org/10.1148/ryai.2021200126> (2021).
- Wong, J., Manderson, T., Abrahamowicz, M., Buckeridge, D. L. & Tamblyn, R. Can hyperparameter tuning improve the performance of a super learner?. *Epidemiology* **30**, 521–531. <https://doi.org/10.1097/EDE.0000000000001027> (2019).

## Acknowledgements

We thank São Paulo Research Foundation (FAPESP) for financial support under the Grant 2019/05723-7 and for the scholarships 2020/00037-5, 2020/07411-0, 2020/01079-3 and 2021/04199-2. We thank the Brazilian Council for Development of Science and Technology (CNPq) for the scholarships 136884/2020-2 and 118670/2019-0. S.K.O. would also like to thank CNPq for Grant PQ 304409/2021-9. The opinions, hypotheses, conclusions or recommendations expressed in this material are solely responsibility of the authors and do not necessarily reflect FAPESP's or CNPq's view. This work was fully supported by MaChiron Ltd and Hospital das Clínicas of Sao Paulo.

## Author contributions

B.A.R. conceived the study, led the HepatIA project and wrote the manuscript. L.C.F., J.M.C.F., L.S.L.N. and M.R.M.S.F. worked in data curation. L.G.R.V. was the project manager and developed the methodology, L.G.G.F. worked in methodology development. A.C.M.C. and J.M.R.S. performed the data analysis and wrote the manuscript. A.S.N. developed the platform for exam acquisition, annotation and data storage. C.C.L., C.H.N., M.M.F., G.G.C. and F.J.C. are supervised the project and facilitated the partnership between Faculty of Medicine-University of Sao Paulo, Hospital das Clinicas of Sao Paulo and MaChiron. M.A.G. and S.K.O. were the advisors of this project, aiding in the conception, manuscript review and funding acquisition.

### Competing interests

A.C.M.C, B.A.R. and L.G.R.V. are cofounders of Machiron Ltd. FMUSP, HCFMUSP and Machiron Ltd. signed a collaboration agreement. The terms of this arrangement have been reviewed and approved by University of São Paulo in accordance with its conflict of interest policies. L.C.F., J.M.C.F., L. S. L. N., M. R. M. S. F., L.G.G.F, J.M.R.S, A.S.N., C.C.L., C.H.N., M.M.F, G.G.C., F. J. C.,M.A.G. and S.K.O. declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to B.A.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022