Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

ELSEVIER

Original Research

An artificial intelligence model predicts the survival of solid tumour patients from imaging and clinical data

Kathryn Schutte [a,*], Fabien Brulport [a], Sana Harguem-Zayani [b],
Jean-Baptiste Schiratti [a], Ridouane Ghermi [a], Paul Jehanno [a],
Alexandre Jaeger [a,e], Talal Alamri [b], Raphaël Naccache [b],
Leila Haddag-Miliani [b], Teresa Orsi [b], Jean-Philippe Lamarque [c],
Isaline Hoferer [b,d], Littisha Lawrance [b,d], Baya Benatsou [b,d],
Imad Bousaid [c], Mikael Azoulay [c], Antoine Verdon [c], François Bidault [b,d],
Corinne Balleyguier [b,d], Victor Aubert [a], Etienne Bendjebbar [a],
Charles Maussion [a], Nicolas Loiseau [a], Benoît Schmauch [a],
Meriem Sefta [a], Gilles Wainrib [a], Thomas Clozel [a], Samy Ammari [b,d],
Nathalie Lassau [b,d]

[a] *Owkin Lab, Owkin, Inc., 10003, New York, NY, USA*
[b] *Department of Imaging, Gustave Roussy, Université Paris Saclay, 94805, Villejuif, France*
[c] *Direction of Digital Transformation & Information Systems, Gustave Roussy, Université Paris-Saclay, 94805, Villejuif, France*
[d] *Biomaps, UMR1281 INSERM, CEA, CNRS, Université Paris-Saclay, 94805, Villejuif, France*
[e] *Calypse Consulting, 75002, Paris, France*

Abstract *Background:* The need for developing new biomarkers is increasing with the emergence of many targeted therapies. Artificial Intelligence (AI) algorithms have shown great promise in the medical imaging field to build predictive models. We developed a prognostic model for solid tumour patients using AI on multimodal data.
Patients and methods: Our retrospective study included examinations of patients with seven different cancer types performed between 2003 and 2017 in 17 different hospitals. Radiologists annotated all metastases on baseline computed tomography (CT) and ultrasound (US) images. Imaging features were extracted using AI models and used along with the patients' and treatments' metadata. A Cox regression was fitted to predict prognosis. Performance was assessed

* *Corresponding author*:
  E-mail address: kathryn.schutte@owkin.com (K. Schutte).

on a left-out test set with 1000 bootstraps.

Results: The model was built on 436 patients and tested on 196 patients (mean age 59, IQR: 51 −6, 411 men out of 616 patients). On the whole, 1147 US images were annotated with lesions delineation, and 632 thorax-abdomen-pelvis CTs (total of 301,975 slices) were fully annotated with a total of 9516 lesions. The developed model reaches an average concordance index of 0.71 (0.67−0.76, 95% CI). Using the median predicted risk as a threshold value, the model is able to significantly (log-rank test P value < 0.001) isolate high-risk patients from low-risk patients (respective median OS of 11 and 31 months) with a hazard ratio of 3.5 (2.4 −5.2, 95% CI).

Conclusion: AI was able to extract prognostic features from imaging data, and along with clinical data, allows an accurate stratification of patients' prognoses.

1. Introduction

Antiangiogenics are widely used to treat solid tumours and improve the outcome of patients by blocking angiogenesis [1,2]. The prescription of antiangiogenic treatments has been approved for many metastatic solid tumours such as kidney, colorectal, lung, liver, GIST and advanced breast cancer, among others. Clinical trials on many other types of cancer are ongoing.

Several preliminary studies on cancer-specific cohorts [3,4] have shown that dynamic contrast-enhanced (DCE-US) measurements of tumour perfusion [5] at early treatment stages of antiangiogenics are associated with standard endpoints. In 2014, the STIC (Soutien aux Techniques Innovantes Coûteuses) study [6] confirmed in a large prospective multicentric cohort that the analysis of DCE-US enables the detection of early response to antiangiogenics for solid tumours. The newly discovered criterion was then added to the International Guidelines and Good Clinical Practice Recommendations for contrast-enhanced ultrasound [7].

However, despite further efforts to find prognostic associations at earlier stages [8], no significant association was found between baseline criteria values and standard endpoints such as progression-free survival (PFS) and overall survival (OS).

Finding a baseline prognostic marker would have a great impact on treatment planning by identifying patients with high and low risks of survival under antiangiogenic treatment.

The use of Artificial intelligence (AI) in the field of medical imaging has recently shown great promise [9,10]. Schmauch et al. [11] used deep learning algorithms on ultrasound images to accurately detect and characterise focal liver lesions. Similarly, P. Blanc-Durand et al. demonstrated that deep learning methods can accurately segment muscular body mass from computed tomography (CT) images to quantify sarcopenia [12]. More recently, Lassau et al. [13] made use of AI algorithms' ability to aggregate information contained in multiple modalities to predict the severity of COVID-19 patients.

However, the quality and quantity of available datasets in the medical field represent a major obstacle to developing robust algorithms.

In the current retrospective multicentric study, we collected and annotated a large cohort of CT and ultrasound (US) images and leveraged deep learning algorithms to develop a multimodal model predictive of response to antiangiogenics for solid tumours.

2. Materials and methods

2.1. Data collection

Our retrospective study uses a cohort collected from 2003 to 2017, which includes solid tumour patients treated with antiangiogenics during different clinical trials of multiple centres that had DCE-US follow-up [14].

Baseline images were taken within 60 days before treatment initiation and 7 days after treatment initiation. Patients and treatment metadata were collected, as well as follow-up data, including PFS and OS.

All patients provided written informed consent. The study was approved by the ethics committee of our institution and was declared to the French Commission Nationale Informatique et Liberté (CNIL MR-004).

All patients consisted of one or two (orthogonal) US images centred on one target lesion and one thorax-abdomen-pelvis CT scan. Expert radiologists drew a delineation of the visible lesion on each US image and annotated all visible lesions on each CT with a bounding box.

In both CT and US images, AI features were extracted from the regions of the lesions.

Finally, additional features related to the tumour burden were extracted from the annotations of all lesions on CT: the number of lesions and the tumour burden volume in the liver, the lungs and elsewhere approximated by an ellipsoid inscribed in the bounding boxes.

Clinical features used in this study included patients' metadata (sex, age, weight), cancer type and treatment

metadata (treatment line, treatment molecule and molecule count).

More details on data annotation and preprocessing can be found in the supplementary material.

2.2. Modeling

All prognosis models are Cox Regression models, implemented in the Lifelines [15] python package.

CT features were aggregated across lesions using max pooling to obtain a single set of features (2048 AI features). AI features were normalised using standard scaling, and principal component analysis (PCA) was used to obtain 10 features per modality. Input features of the multimodal model consist of the concatenation of only five principal components per imaging modality.

2.3. Statistical analysis

A test set of 196 patients obtained from a train-test split stratified on patients and cancer types was left out of the study for validation purposes.

Models were compared using the concordance index (c-index) on 1000 sets of 196 patients drawn from the test set with replacement (bootstrap method). The performance of a model is considered to be significantly superior to another model if it is superior on more than 95% of the 1000 comparisons. To evaluate the performance of models to stratify patients, we used the median predicted risk value as a threshold. Hazard ratios (HR) were compared using the same method used for the c-index comparisons. Finally, a log-rank test was performed on the test set and Kaplan−Meier curves were drawn.

A subgroup analysis was performed on the two groups of high-risk and low-risk patients identified by the model. Continuous variables for each subgroup were compared using a T-test, and binary variables were compared using a proportions Z-test. P values were corrected for the number of tests and considered significant if <0.05.

3. Results

3.1. Cohort description

This study included a total of 632 patients from the initial cohort of 1034 patients. A first selection was made on the cancer type to include only sufficiently represented cancer types: primary hepatocellular carcinoma (HCC), metastatic renal cell carcinoma (RCC), colorectal carcinoma (CRC), gastrointestinal stromal tumour (GIST), melanoma, breast cancer or sarcoma, which resulted in the

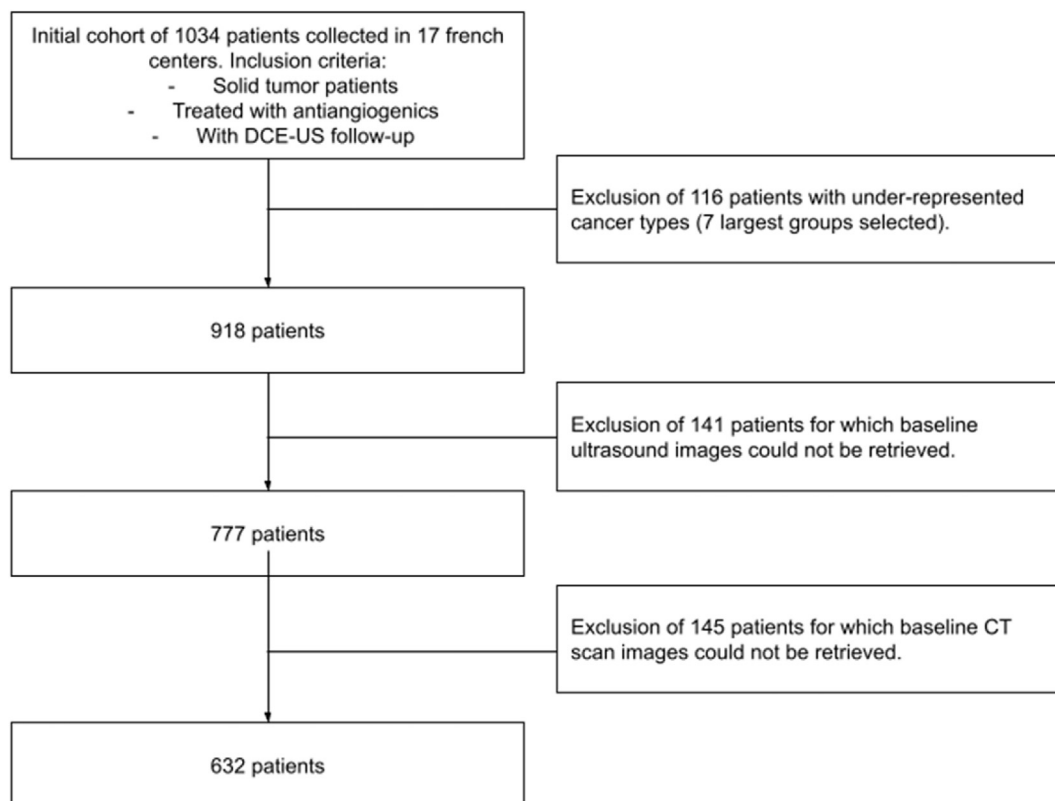Fig. 1. **Flow chart of study inclusion and exclusion criteria**. The seven most-represented cancer types were selected from the original cohort: hepatocellular carcinoma, renal cell carcinoma, colorectal carcinoma, gastrointestinal stromal tumour, melanoma, breast cancer and sarcoma. Patients for whom baseline images were taken within 60 days before treatment initiation and 7 days after treatment initiation were selected.

exclusion of 116 patients. Then, we excluded 286 patients for which no baseline US or baseline CT images could be retrieved. The inclusion and exclusion criteria are described in the flow chart in Fig. 1.

The median PFS of the cohort is 9 months, and the median OS is 14 months. During the observed period, 480 (75.9% of patients) death events occurred. Clinical characteristics of patients are described in Table 1: 411 men and a median age of 59 years (IQR: 51–67).

The annotation process resulted in 632 fully annotated patients; 1147 US images were annotated with lesions delineation, and 632 CTs (301,975 slices) were fully annotated with up to 150 lesions per patient for a total of 9516 annotated lesions.

3.2. A multimodal prognostic model

Our multimodal prognostic model (PULS-AI) was fitted to the concatenation of four sets of variables: clinical variables, AI features extracted from the US target lesion, AI features extracted from all lesions on CT images and tumour burden features extracted from CT annotations (Fig. 2). This model reaches an average c-index of 0.71 (0.67–0.76, 95% CI). The developed model has an HR of 3.5 (2.4–5.2, 95% CI) and the Kaplan–Meier analysis, as shown in Fig. 2, demonstrates the ability of this model to significantly (log-rank test P value < 0.001) stratify patients into a high-risk group (median OS of 11 months) and a low-risk group (median OS of 31 months).

3.3. High-risk and low-risk subgroups analysis

A subgroup analysis was performed on the high-risk patients and the low-risk patients as predicted by the PULS-AI model. Variables that significantly differ between the two subgroups are described in Table 2. Patients in the high-risk group had a significantly (T-test P value <0.001) lower weight (median 70 kg, IQR: 60–72) than patients in the low-risk group (median 70 kg, IQR: 70–82). The number of lesions significantly differed as well, with a median of 13 lesions (IQR: 7–25) in the high-risk group and seven lesions (IQR: 3–13) in the low-risk group (T-test P value <0.001). Similarly, the tumour volume in the liver and the lungs were higher in the high-risk group than in the low-risk group: a median of 124 cm^3 (IQR: 4–529) versus 15 cm^3 (IQR: 0–109) of liver tumour (T-test P value = 0.01) and a median of 1 cm^3 (IQR: 0–33) versus 0 cm^3 (IQR: 0–5) of lung tumour (T-test P value = 0.02). Regarding the AI imaging features, the first and third principal components of the CT features and the first principal component of the US features differed significantly between the two groups (T-test P values = 0.01). Unsurprisingly, the distribution of cancer types and treatments significantly differed among the two subgroups, as described in Table 2.

Table 1

Cohort description and association between variables and overall survival.

Variable	N (%) Total = 632	HR	P value association with survival
Women	221 (34.9)	1.1 (0.9–1.3)	–
Age, median (IQR), y	59 (51–67)	1.0 (0.9–1.1)	–
Weight, median (IQR), kg	71 (60–82) *mv: 183*	0.8 (0.8–0.9)	<0.001
Treatment line	1: 305 (48.3)	1.3 (1.2–1.4)	<0.001
	2: 167 (26.3)		
	3: 72 (11.5)		
	4–10: 87 (13.7) *mv: 1*		
Treatment molecules count	1: 507 (80.0)	1.0 (1.0–1.1)	–
	2: 98 (15.8)		
	3: 24 (3.8)		
	4: 3 (0.5)		
Treatment group			
TKI VEGFR	354 (56.2)	1.5 (1.2–1.7)	<0.001
Mab anti-VEGFR	151 (23.8)	0.8 (0.7–1.0)	–
c-KIT	59 (9.3)	0.3 (0.2–0.5)	<0.001
Others	68 (10.7)	2.0 (1.5–2.7)	<0.001
Cancer			
RCC	253 (39.9)	0.9 (0.7–1.1)	–
CRC	97 (15.3)	1.2 (0.9–1.5)	–
CHC	93 (14.7)	1.9 (1.5–2.4)	<0.001
GIST	70 (11.0)	0.4 (0.3–0.5)	<0.001
Melanoma	52 (8.5)	2.3 (1.7–3.2)	<0.001
Breast	37 (5.8)	0.8 (0.5–1.2)	–
Sarcoma	30 (4.7)	1.6 (1.0–2.4)	0.04

Variables are described as median (IQR) or as N (%). Associations with OS are evaluated with a Cox Regression and reported with hazard ratios (HR) and P values when significant (<0.05). For continuous variables, HRs are computed for an increase of one standard deviation of the continuous variable. *mv*: missing values, *y*: years, *kg*: kilograms.

3.4. Ablation study: evaluation of US and CT AI features predictive value

As target lesions were assessed on US and CT images, we wanted to evaluate the signal contained in both images separately and investigate whether both images contain complementary information. To that end, we fitted a first model (AI-US-Target) on AI features extracted from the US images of target lesions, a second model (AI-CT-Target) on AI features extracted from the CT images of target lesions (excluding all other lesions) and finally a third model (AI-CTUS-Target) on the combination of AI features extracted from both image types. As described in Table 3, The AI-CT-Target model does not perform significantly better than a random predictor, with a c-index of 0.53 (0.48–0.59, 95% CI). The Kaplan–Meier analysis shows that AI-US-Target can significantly identify high-risk patients and low-risk patients with a log-rank test P value of 0.01 and an HR of 1.7 (1.2–2.4). The given results show that US and CT imaging do not hold complementary information as the performance obtained when combining
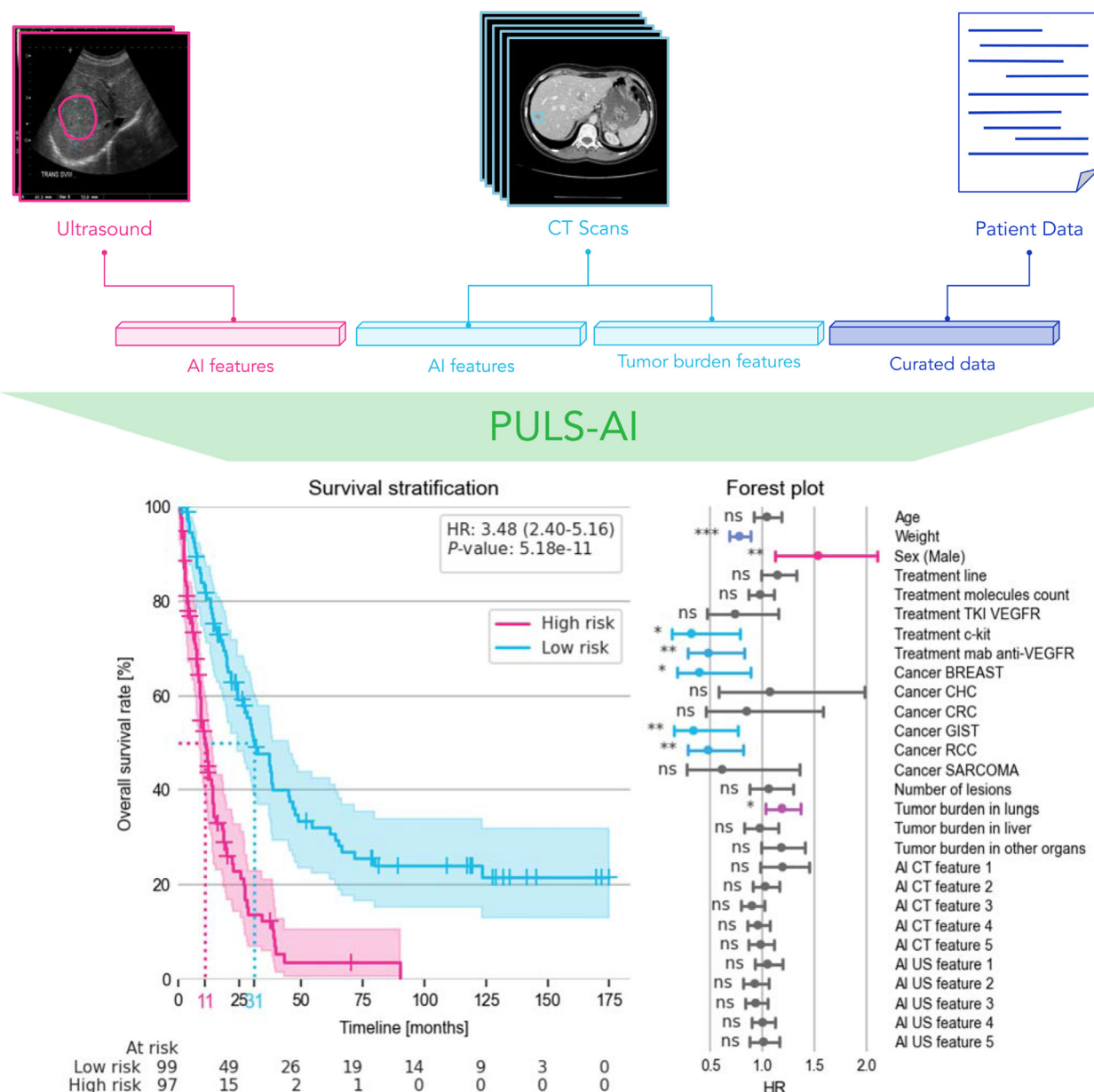
Fig. 2. **PULS-AI layout and results.** AI features were extracted from ultrasound images and CT scan images using expert lesions annotations (bounding boxes). Features descriptive of the tumour burden were computed using the lesions annotations on CT scans (total amount of lesions and the tumour burden volume per organ). Finally, patient, cancer and treatment data were curated to be included in the model. A Cox regression model is fitted to the concatenated feature sets and predicts the mortality risk for each patient. High-risk patients and low-risk patients are identified using the median risk score of PULS-AI as a threshold value. The model's predictions are prognostic scores (HR = 3.48, 95% CI: 2.38–5.17, log-rank test P value < 0.001). Kaplan−Meier curves are drawn with a 95% confidence interval for the two groups of patients. A forest plot shows the hazard ratios with a 95% confidence interval for the variables of the multivariate Cox regression (PULS-AI). $*{*}{*}$: $P < 0.001$, $*{*}$: $P < 0.01$, $*$: $P < 0.05$, ns: not significant.

both images is not increased with a c-index of 0.57 (0.52−0.62) for AI-US-Target and 0.56 (0.51−0.62) for AI-CTUS-Target.

3.5. Ablation study: evaluation of the tumour burden predictive value

Next, we wanted to measure the predictive power of tumour burden features extracted from CT images. To that end, we fitted a model to the tumour burden features alone (Cox-TB). Cox-TB model reaches a c-index of 0.63 (0.58−0.68, 95% CI) and is able to significantly stratify patients into two risk groups (log-rank test P value < 0.001) with an HR of 2.32 (1.64−3.36, 95% CI) as shown in Table 3. Interestingly, the performance of Cox-TB is similar to the performance of AI-CT+, which may indicate that the most valuable information derived from CT images lies in the tumour burden assessment.

Table 2
Subgroups analysis of the high-risk and low-risk patients defined by the PULS-AI model.

Feature	T-test P value	High-risk Group median (IQR)	Low-risk Group median (IQR)
AI CT 0	<0.001	8.4 (−12.3−32.2)	−11.1 (−30.0−10.7)
AI CT 2	<0.001	−1.8 (−6.9−3.3)	1.0 (−4.2−6.7)
AI US 0	0.01	−3.1 (−12.2−10.9)	−5.5 (−12.7−3.4)
Weight (kg)	<0.001	1	70 (70−82)
Lesions count	<0.001	13 (7−25)	6 (3−13)
Lung Volume (cm3)	0.02	1 (0−33)	0 (0−5)
Liver Volume (cm3)	0.01	124 (4−529)	15 (0−109)
Feature	**Proportion Z-test P value**	**High risk (%)**	**Low risk (%)**
Treatment TKI-VEGFR	0.01	63.3	48.7
Treatment C-KIT	<0.001	1.0	17.7
Treatment mab Anti-VEGFR	<0.001	15.8	32.0
Cancer CHC	<0.001	25.3	4.1
Cancer CRC	0.02	20.3	10.4
Cancer GIST	<0.001	1.6	20.6
Cancer RCC	<0.001	27.5	52.5
Cancer Sarcoma	0.01	7.9	1.6

The table describes variables that are statistically different between the two groups. Continuous variables were tested using a T-test and are described by the median and IQR. Binary variables were tested using a Proportion Z-test and are described by the corresponding percentage of patients within each group. P values were corrected for multiple testing and considered significant if < 0.05.

3.6. Ablation study: comparison of single-modality models vs. multimodality models

Finally, we wanted to compare the prognostic value of imaging data and clinical data and assess their complementarity. To that end, we compared three models: a model fitted to the imaging features that include AI features extracted from US images of the target lesions, AI features extracted from CT images of all lesions, and the handcrafted tumour burden features (AI-USCT+), a model fitted to the clinical variables alone (Clin-Cox) and the multimodal prognosis model (PULS-AI) introduced earlier. The three models shown in Fig. 3 are prognostic with an HR of 2.1 (1.5−3.1, 95% CI), 3.0 (2.1−4.4, 95% CI) and 3.5 (2.4−5.2, 95% CI) respectively, and significant log-rank test P values (<0.001). However, PULS-AI significantly outperforms (P values < 0.001) AI-USCT+ and Clin-Cox in terms of c-index, with 0.71 (0.67−0.76, 95% CI) points compared to 0.61

(0.56−0.66, 95% CI) and 0.68 (0.63−0.72, 95% CI) points respectively.

Statistical comparisons of models are presented in Supplementary eTable 2.

4. Discussion

Treatment decisions are crucial steps in a patient's therapeutic pathway and highly impact a patient's prognosis. As a patient progresses into successive lines of treatments, therapeutic options and guidelines get fewer, which leaves medical staff to their own experience in deciding on the subsequent treatment. Developing reliable and reproducible prognostic markers are key to improving patients' outcomes.
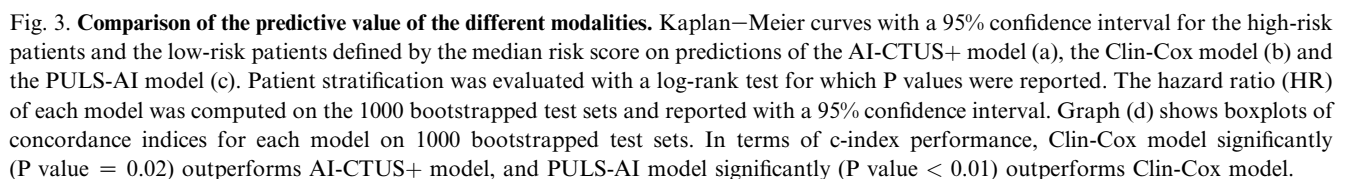
For the purpose of this study, we built a unique cohort of extensively and manually annotated baseline radiology examinations from 632 patients collected from 17 different French centres.

Table 3
Ablation study.

Model	C-index (95% CI)	HR (95% CI)	Log-rank test P value	Clinical features	AI US features	AI CT features	CT tumour burden features
Puls-AI	0.71 (0.67−0.76)	3.5 (2.4−5.2)	<0.001	X	X	X	X
AI-US target	0.57 (0.52−0.62)	1.7 (1.2−2.4)	0.01		X		
AI-CT Target	0.53 (0.48−0.59)	1.5 (1.1−2.0)	0.06			target only	
AI-US-CT Target	0.56 (0.51−0.62)	1.7 (1.2−2.4)	0.01		X	target only	
AI-CT+	0.61 (0.56−0.66)	1.9 (1.3−2.7)	<0.001			X	X
TB-Cox	0.63 (0.58−0.68)	2.3 (1.6−3.4)	<0.001				X
AI-US-CT+	0.61 (0.56−0.67)	2.1 (1.5−3.1)	<0.001		X	X	X
Clin-Cox	0.68 (0.63−0.72)	3.0 (2.1−4.4)	<0.001	X			

The performance of each model is reported in terms of C-index and HR with a 95% confidence interval. Additionally, P values of the log-rank tests are reported. For each model, the input features used are specified.

Fig. 3. **Comparison of the predictive value of the different modalities.** Kaplan−Meier curves with a 95% confidence interval for the high-risk patients and the low-risk patients defined by the median risk score on predictions of the AI-CTUS+ model (a), the Clin-Cox model (b) and the PULS-AI model (c). Patient stratification was evaluated with a log-rank test for which P values were reported. The hazard ratio (HR) of each model was computed on the 1000 bootstrapped test sets and reported with a 95% confidence interval. Graph (d) shows boxplots of concordance indices for each model on 1000 bootstrapped test sets. In terms of c-index performance, Clin-Cox model significantly (P value = 0.02) outperforms AI-CTUS+ model, and PULS-AI model significantly (P value < 0.01) outperforms Clin-Cox model.

In this study, we developed an AI-based multimodal model to predict the prognosis of solid tumour patients treated with antiangiogenics.

Our model aggregates variables extracted from US images, CT scan images and clinical data to compute a prognostic score (HR = 3.5 (2.4−5.2, 95% CI)) that allows stratifying patients into two subgroups of high risks (median OS of 11 months) and low risks (median OS of 31 months) of death.

We ran a subgroup analysis to identify variables that significantly differed between high-risk patients and low-risk patients. Imaging data reveals significant differences in three AI-extracted features, as well as in tumour burden features with a higher number of metastases and a higher tumoural volume in lungs and liver for the high-risk group of patients.

Additionally, we have shown that AI features extracted from a single lesion on US images alone hold a prognostic value that is sufficient to significantly stratify patients.

Similarly, we have shown that the tumour burden features extracted from the CT alone allow for building a solid prognostic marker. This result suggests that most of the predictive value contained in CT images is derived from the assessment of the complete tumour burden.

Finally, we have demonstrated that a model on the combination of both imaging data and clinical data yields significantly better results than building models in every single modality separately, proving the complementarity

of the different data modalities. This result supports the necessity of building composite biomarkers from multiple sources of data.

Our study has several limitations. First, we were not able to constitute an external validation cohort to validate our results. Although the size and the multicentric character of the cohort should allow for robust results and mitigate the false discovery risk often discussed regarding quantitative imaging biomarkers [16−18], the developed model requires further validation on an external cohort before any clinical adoption, in accordance with the imaging biomarker roadmap for cancer research [19]. Second, the clinical data collected did not allow us to compare PULS-AI's performance to existing prognostic scores, such as the IMDC for metastatic RCC [20] or the Glasgow prognostic score for CRC [21] and CHC [22].

Finally, the applicability of PULS-AI is limited today as it requires the annotation of all visible lesions on patients' CT scans, which is time consuming and thus not required in routine. However, we have shown the important benefit of assessing the complete tumour burden, and we believe that this task can easily be—and will soon be—automated by an AI algorithm.

Further investigations may focus on the medical interpretability of the newly developed model. Many techniques have been recently developed to allow a better understanding of deep learning-based predictions. One such technique, the GradCam method, allows visualising areas in a given image that impact the prediction. A more recent approach uses Generative models [23] to show how specific features of an image have to change in order to produce different predictions [24].

The results of this prognostic study show that AI algorithms are able to extract relevant information from radiology images and aggregate data from multiple modalities to build powerful prognostic tools. Such tools could provide assistance to oncology clinicians in therapeutic decision-making.

Financial support

Author contributions

Conceptualisation: Kathryn Schutte, Paul Jehanno, Samy Ammari, Nathalie Lassau, Victor Aubert, Etienne Bendjebbar, Charles Maussion, Meriem Sefta.

Methodology: Kathryn Schutte, Paul Jehanno, Samy Ammari, Nathalie Lassau, Charles Maussion, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi.

Software and investigations: Kathryn Schutte, Paul Jehanno, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi.

Formal analysis: Kathryn Schutte, Paul Jehanno, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi, Nicolas Loiseau.

Resources: Sana Harguem-Zayani, Alexandre Jaeger, Talal Alamri, Raphaël Naccache, Leila Haddag-Miliani, Teresa Orsi, Jean-Philippe Lamarque, Isaline Hoferer, Littisha Lawrance, Baya Benatsou, Imad Bousaid, Mikael Azoulay, Antoine Verdon, François Bidault, Corinne Balleyguier, Gilles Wainrib, Thomas Clozel, Samy Ammari, Nathalie Lassau.

Data curation: Sana Harguem-Zayani, Alexandre Jaeger, Talal Alamri, Raphaël Naccache, Leila Haddag-Miliani, Teresa Orsi, Jean-Philippe Lamarque, Isaline Hoferer, Littisha Lawrance, Baya Benatsou, Imad Bousaid, Mikael Azoulay, Antoine Verdon, François Bidault, Corinne Balleyguier, Gilles Wainrib, Thomas Clozel, Samy Ammari, Nathalie Lassau.

Writing - original draft: Kathryn Schutte, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi, Samy Ammari, Nathalie Lassau.

Writing - review and editing: Kathryn Schutte, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi, Nicolas Loiseau, Samy Ammari, Nathalie Lassau, Charles Maussion, Benoit Schmauch.

Supervision and project administration: Kathryn Schutte, Nathalie Lassau, Charles Maussion, Gilles Wainrib, Thomas Clozel.

Conflict of interest statement

The authors declare the following financial interests/personal relationships, which may be considered as potential competing interests: The authors declare the following competing interests: KS, FB, JBS, RG, PJ, AJ, VA, EB, CM, NL, BS, MS are employed by Owkin. GW and TC are co-founders of Owkin. NL reports to be a Speaker at Jazz Pharmaceuticals. NL reports a grant from Guerbet.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2022.06.055.

References

[1] Demetri GD, Reichardt P, Kang Y-K, et al. Efficacy and safety of regorafenib for advanced gastrointestinal stromal tumours after failure of imatinib and sunitinib (GRID): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet Lond Engl 2013;381(9863):295−302. https://doi.org/10.1016/S0140-6736(12)61857-1.

[2] Escudier B, Eisen T, Stadler WM, et al. Sorafenib in advanced clear-cell renal-cell carcinoma. N Engl J Med 2007;356(2):125−34. https://doi.org/10.1056/NEJMoa060655.

[3] Lassau N, Koscielny S, Albiges L, et al. Metastatic renal cell carcinoma treated with sunitinib: early evaluation of treatment response using dynamic contrast-enhanced ultrasonography. Clin

Cancer Res Off J Am Assoc Cancer Res 2010;16(4):1216−25. https://doi.org/10.1158/1078-0432.CCR-09-2175.

[4] Lassau N, Lamuraglia M, Vanel D, et al. Doppler US with perfusion software and contrast medium injection in the early evaluation of isolated limb perfusion of limb sarcomas: prospective study of 49 cases. Ann Oncol Off J Eur Soc Med Oncol 2005; 16(7):1054−60. https://doi.org/10.1093/annonc/mdi214.

[5] Dietrich CF, Averkiou MA, Correas J-M, Lassau N, Leen E, Piscaglia F. An EFSUMB introduction into Dynamic Contrast-Enhanced Ultrasound (DCE-US) for quantification of tumour perfusion. Ultraschall Med Stuttg Ger 1980 2012;33(4):344−51. https://doi.org/10.1055/s-0032-1313026.

[6] Lassau N, Bonastre J, Kind M, et al. Validation of dynamic contrast-enhanced ultrasound in predicting outcomes of anti-angiogenic therapy for solid tumors. Invest Radiol 2014;49(12): 794−800. https://doi.org/10.1097/RLI.0000000000000085.

[7] Sidhu PS, Cantisani V, Dietrich CF, et al. The EFSUMB guidelines and Recommendations for the clinical Practice of contrast-enhanced ultrasound (CEUS) in non-hepatic applications: update 2017 (long version). Ultraschall Med Stuttg Ger 1980 2018; 39(2):e2−44. https://doi.org/10.1055/a-0586-1107.

[8] Lassau N, Coiffier B, Kind M, et al. Selection of an early biomarker for vascular normalization using dynamic contrast-enhanced ultrasonography to predict outcomes of metastatic patients treated with bevacizumab. Ann Oncol Off J Eur Soc Med Oncol 2016; 27(10):1922−8. https://doi.org/10.1093/annonc/mdw280.

[9] Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Health 2020;2(9): e486−8. https://doi.org/10.1016/S2589-7500(20)30160-6.

[10] Lee LIT, Kanthasamy S, Ayyalaraju RS, Ganatra R. The current state of artificial intelligence in medical imaging and nuclear medicine. BJR|Open 2019;1(1):20190037. https://doi.org/10.1259/bjro.20190037.

[11] Schmauch B, Herent P, Jehanno P, et al. Diagnosis of focal liver lesions from ultrasound using deep learning. Diagn Interv Imag 2019;100(4):227−33. https://doi.org/10.1016/j.diii.2019.02.009.

[12] Blanc-Durand P, Schiratti J-B, Schutte K, et al. Abdominal musculature segmentation and surface prediction from CT using deep learning for sarcopenia assessment. Diagn Interv Imag 2020; 101(12):789−94. https://doi.org/10.1016/j.diii.2020.04.011.

[13] Lassau N, Ammari S, Chouzenoux E, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. Nat Commun 2021;12(1):634. https://doi.org/10.1038/s41467-020-20657-4.

[14] Lassau N, Chapotot L, Benatsou B, et al. Standardization of dynamic contrast-enhanced ultrasound for the evaluation of antiangiogenic therapies: the French multicenter Support for Innovative and Expensive Techniques Study. Invest Radiol 2012;47(12):711−6. https://doi.org/10.1097/RLI.0b013e31826dc255.

[15] Davidson-Pilon C, Kalderstam J, Jacobson N, et al. CamDavidsonPilon/lifelines: 0.26.0. Zenodo 2021. https://doi.org/10.5281/zenodo.4816284.

[16] Fournier L, Costaridou L, Bidaut L, et al. Correction to: incorporating radiomics into clinical trials: expert consensus endorsed by the European Society of Radiology on considerations for data-driven compared to biologically driven quantitative biomarkers. Eur Radiol 2021;31(8):6408−9. https://doi.org/10.1007/s00330-021-07721-3.

[17] Espinasse M, Pitre-Champagnat S, Charmettant B, et al. CT texture analysis challenges: influence of acquisition and reconstruction parameters: a comprehensive review. Diagn Basel Switz 2020;10(5):E258. https://doi.org/10.3390/diagnostics10050-258.

[18] Caramella C, Allorant A, Orlhac F, et al. Can we trust the calculation of texture indices of CT images? A phantom study. Med Phys 2018;45(4):1529−36. https://doi.org/10.1002/mp.12809.

[19] O'Connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017; 14(3):169−86. https://doi.org/10.1038/nrclinonc.2016.162.

[20] Prognostic Factors for Overall Survival in Patients With Metastatic Renal Cell Carcinoma Treated With Vascular Endothelial Growth Factor−Targeted Agents: Results From a Large, Multicenter Study | Journal of Clinical Oncology n.d. https://ascopubs.org/doi/10.1200/JCO.2008.21.4809 (accessed November 3, 2021).

[21] Nozoe T, Matono R, Ijichi H, Ohga T, Ezaki T. Glasgow prognostic score (GPS) can Be a useful indicator to determine prognosis of patients with colorectal carcinoma. Int Surg 2014;99(5): 512−7. https://doi.org/10.9738/INTSURG-D-13-00118.1.

[22] Kinoshita A, Onoda H, Imai N, et al. The Glasgow Prognostic Score, an inflammation based prognostic score, predicts survival in patients with hepatocellular carcinoma. BMC Cancer 2013; 13(1):52. https://doi.org/10.1186/1471-2407-13-52.

[23] Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. 2019. ArXiv181204948 Cs Stat.

[24] Schutte K, Moindrot O, Hérent P, Schiratti J-B, Jégou S. Using StyleGAN for visual interpretability of deep learning models on medical images. 2021. ArXiv210107563 Cs Eess.