GEP-NET radiomics: a systematic review and radiomics quality score assessment

European Radiology
https://doi.org/10.1007/s00330-022-08996-w

ONCOLOGY

GEP-NET radiomics: a systematic review and radiomics quality score assessment

Femke C. R. Staal[1,2,3] · Else A. Aalbersberg[3,4] · Daphne van der Velden[1] · Erica A. Wilthagen[5] · Margot E. T. Tesselaar[3,6] · Regina G. H. Beets-Tan[1,2,7] · Monique Maas[1]

Received: 21 February 2022 / Revised: 25 May 2022 / Accepted: 26 June 2022

Abstract

Objective The number of radiomics studies in gastroenteropancreatic neuroendocrine tumours (GEP-NETs) is rapidly increasing. This systematic review aims to provide an overview of the available evidence of radiomics for clinical outcome measures in GEP-NETs, to understand which applications hold the most promise and which areas lack evidence.

Methods PubMed, Embase, and Wiley/Cochrane Library databases were searched and a forward and backward reference check of the identified studies was executed. Inclusion criteria were (1) patients with GEP-NETs and (2) radiomics analysis on CT, MRI or PET. Two reviewers independently agreed on eligibility and assessed methodological quality with the radiomics quality score (RQS) and extracted outcome data.

Results In total, 1364 unique studies were identified and 45 were included for analysis. Most studies focused on GEP-NET grade and differential diagnosis of GEP-NETs from other neoplasms, while only a minority analysed treatment response or long-term outcomes. Several studies were able to predict tumour grade or to differentiate GEP-NETs from other lesions with a good performance (AUCs 0.74–0.96 and AUCs 0.80–0.99, respectively). Only one study developed a model to predict recurrence in pancreas NETs (AUC 0.77). The included studies reached a mean RQS of 18%.

Conclusion Although radiomics for GEP-NETs is still a relatively new area, some promising models have been developed. Future research should focus on developing robust models for clinically relevant aims such as prediction of response or long-term outcome in GEP-NET, since evidence for these aims is still scarce.

Key Points

- *The majority of radiomics studies in gastroenteropancreatic neuroendocrine tumours is of low quality.*
- *Most evidence for radiomics is available for the identification of tumour grade or differentiation of gastroenteropancreatic neuroendocrine tumours from other neoplasms.*
- *Radiomics for the prediction of response or long-term outcome in gastroenteropancreatic neuroendocrine tumours warrants further research.*

Keywords Neuroendocrine tumors · Gastrointestinal neoplasms · Artificial intelligence · Machine learning

✉ Monique Maas
m.maas@nki.nl

[1] Department of Radiology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[2] GROW School for Oncology and Reproduction, Maastricht University Medical Centre, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands

[3] The Netherlands Cancer Institute/University Medical Center Utrecht Center for Neuroendocrine Tumors, ENETS Center of Excellence, Amsterdam/Utrecht, The Netherlands

[4] Department of Nuclear Medicine, The Netherlands Cancer Institute Amsterdam, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[5] Scientific Information Service, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[6] Department of Medical Oncology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

[7] Faculty of Health Sciences, University of Southern Denmark, J. B. Winsløws Vej 19, 3, 5000 Odense, Denmark

Published online: 26 July 2022

Abbreviations

AP	Arterial phase
DL	Deep learning
G1/2/3	Grade 1/2/3
GEP-NET	Gastroenteropancreatic neuroendocrine tumour
IBSI	Image Biomarker Standardisation Initiative
LASSO	Least absolute shrinkage and selection operator
LR	Logistic regression
ML	Machine learning
NEC	Neuroendocrine carcinoma
PDAC	Pancreatic ductal adenocarcinoma
PFS	Progression-free survival
pNETs	Pancreatic neuroendocrine tumour
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-analysis
PVP	Portal venous phase
RF	Random forest
RFS	Recurrence-free survival
ROI	Region of interest
RQS	Radiomics quality score
SVM	Support vector machine
WHO	World Health Organization

Introduction

During the past decades, it has been established that tumours are heterogeneous entities [1] and it is widely accepted that this heterogeneity has implications for tumour development, treatment outcome and survival [2]. Gastroenteropancreatic neuroendocrine tumours (GEP-NETs) are a rare group of heterogenous tumours with neuro-endocrine differentiation within the gastrointestinal tract or pancreas [3], excluding the poorly differentiated neuroendocrine carcinoma (NEC). Even with available biomarkers and imaging, it remains a challenge to predict the clinical course of an individual patient and select the optimal treatment, because of the heterogeneity between and within tumours [4, 5].

Over the past years, it has been established that medical imaging contains more data than is visible to the naked eye [6, 7] and can be converted to innumerable features and thereby quantify tissue heterogeneity [6, 8, 9]. This technique, radiomics, can describe the relationship between the intensity and position of voxels within an image. Promising results have been achieved for diagnosis, response assessment and prediction of long-term outcome in several tumour types [9, 10]. For GEP-NETs, the potential of radiomics for several aims has been investigated: predicting tumour grade, distinguishing NET from other tumours, and prediction of response and long-term outcomes. However, radiomics has only recently been introduced to GEP-NETs and the number of studies is rapidly increasing, yet with conflicting results. Hence, it is unclear which specific applications of radiomics in the field of GEP-NETs hold the most promise and what areas lack evidence. The radiomics quality score (RQS) is a tool that has been developed specifically to assess the methodological quality of radiomics studies and has not yet been used in GEP-NETs [8, 11, 12]. This systematic review aims to provide an overview of the available literature regarding the use of radiomics in GEP-NETs based on the main aims and to identify promising research directions for future radiomics studies.

Methods and materials

Search strategy

This study was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement [13]. The review protocol is available through PROSPERO (CRD42021226844). The search strategy was conducted by a medical information specialist (E.W.). PubMed, Embase and Scopus were searched until August 2021 (Fig. 1). The following terms, including synonyms and closely related words, were used: 'GEP-NET' AND 'radiomics'. A detailed search strategy is described in the supplement. Citations and references of eligible studies were searched using Scopus to identify further studies for inclusion, until no more new eligible articles were identified.

Study selection

Two reviewers independently reviewed titles and abstracts for eligibility. The first reviewer (F.S.) reviewed all studies and the role of the second reviewer was fulfilled by two reviewers (D.v.d.V. or E.A.). All radiomics analyses were considered for inclusion (i.e. texture analysis, machine learning and deep learning [DL] methods). Articles that met the following criteria were included: (1) patients with GEP-NETs, and (2) radiomics analysis on computed tomography (CT), magnetic resonance imaging (MRI) or positron emission tomography (PET). Reviews, case series ($n < 10$), letters to the editor, conference abstracts and studies with no English text were excluded. Disagreements were resolved by discussion and consensus.

Data collection

Two reviewers independently extracted from the included studies, using a pre-defined data extraction form: study population, clinical outcome, primary tumour, intervention, imaging modality, reference standard, region of interest (ROI), details about the radiomics workflow (including feature extraction, selection, and statistical analysis) and most relevant results. Disagreements were resolved by discussion and consensus. Results were grouped according to three categories based on the main aims: (1) predicting tumour grade, (2)

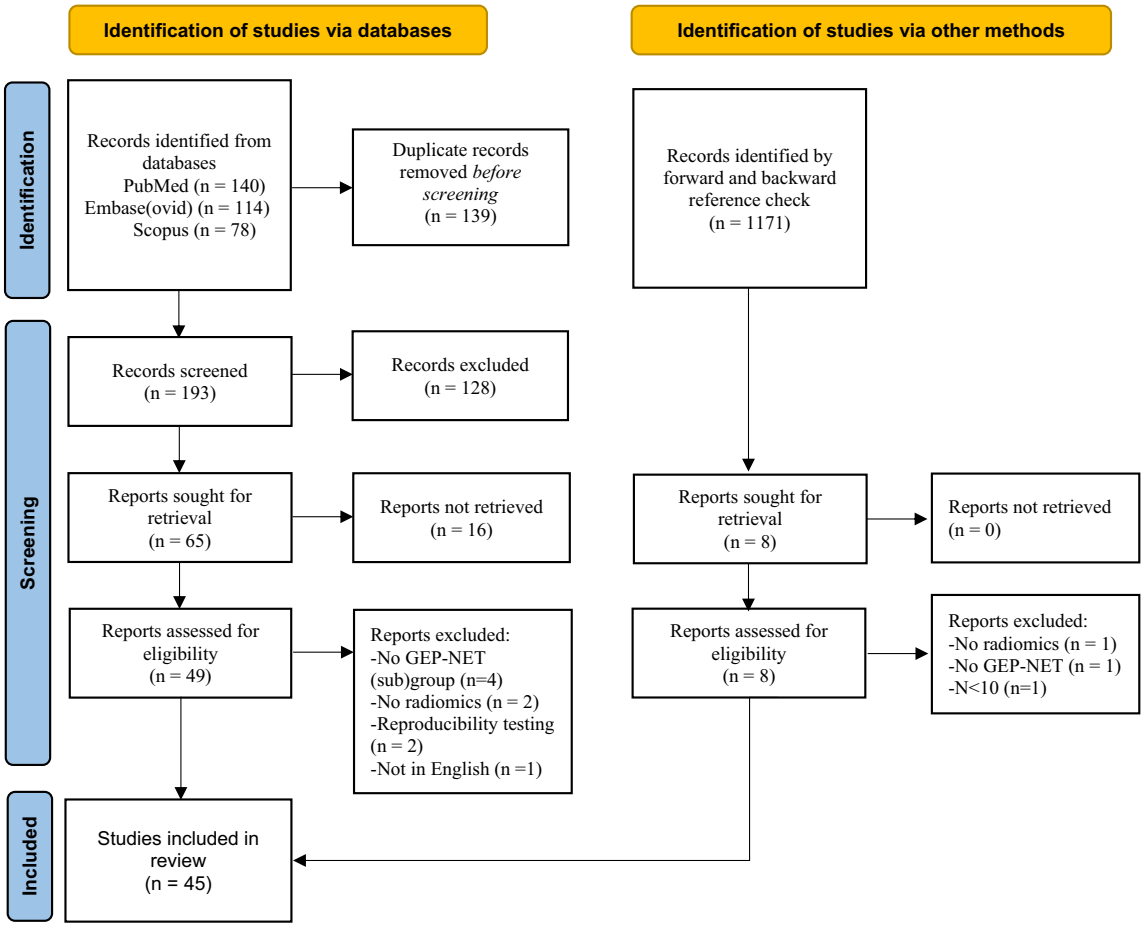

Fig. 1 PRISMA flowchart

distinguishing GEP-NETs from other tumours and (3) response and long-term outcome measures.

Quality assessment

The risk of bias and methodological quality of the studies were assessed independently by two reviewers, using the RQS [8]. This tool was developed to assess the quality of radiomics research. The RQS has 16 components that are rated resulting in a total score ranging from –8 to 36, which is then converted into a percentage score (0–100%). Since the RQS and its components were not developed for DL methods, these were excluded from the RQS evaluation. Disagreements between reviewers were resolved by discussion until consensus.

Statistical analyses

Subgroup analyses were performed to assess the differences in total RQS according to the study aim (predicting tumour grade, distinguishing NET from other tumours and response and long-term outcome measures) and imaging modality (CT, MRI or PET), using the Kruskal-Wallis test. If a study evaluated multiple aims, it was taken into account for each aim. Statistical analyses were performed with Statistical Package for the Social Sciences (SPSS, v27.0). A p value < 0.05 was considered statistically significant.

Results

Study selection

The initial search identified 193 relevant studies; 128 were excluded based on title and abstract screening. Full-text versions of the remaining 65 manuscripts were reviewed and 25 studies were excluded. The backward- and forward reference check of the 40 included studies identified another 5 studies for inclusion. Finally, 45 studies were included for analysis (Fig. 1).

Included studies

The study characteristics, details about radiomics workflow and main results are summarised in Tables 1 and 2. The included studies were published between 2015 and 2021 (Fig.

Table 1 Overview of included studies

Authors	Aim	# Patients			Primary tumour (ROI)	Modality (sequence/phase)	Outcome	Reference standard	Intervention	Main results/conclusion	RQS
		GEP-NET	Other	Total							
Azoulay (2020) [15]	Differentiate pNET from pNEC	14	23	37	pNET/pNEC (pancreas)	CT (AP and PVP)	G3 vs NEC	Pathology	-	Several first-order radiomics features can differentiate between pNET and pNEC	0%
Beleu (2020) [16]	Predict tumour grade	56	-	56	pNET (liver metastases)	CT (AP or PVP)	G1/2 vs G3	Pathology	-	First-order features can predict tumour grade better than shape, margin or calcifications	0%
Benedetti (2021) [17]	Predict tumour grade	39	-	39	pNET (pancreas)	CT (AP and nce)	G1 vs G2/3	Pathology	-	Only asphericity can predict tumour grade	0%
Bevilacqua (2021) [18]	Predict tumour grade	51	-	51	pNET (pancreas)	[^{68}Ga]Ga-DOTANOC PET/CT	G1 vs G2	Pathology	-	A radiomics score with both radiomics and non-radiomics features has the same performance to predict tumour grade than a radiomics score with only radiomics features	36%
Bian (2020) [19]	Predict tumour grade	102	-	102	pNET (pancreas)	CT (PVP)	G1 vs G2	Pathology	-	A CT-based radiomics score is able to predict pNET tumour grade	50%
Bian (2020) [20]	Predict tumour grade	157	-	157	pNET (pancreas)	MRI (T1, AP, PVP)	G1 vs G2/3	Pathology	-	An MRI-based radiomics score is able to predict pNET tumour grade	50%
Bian (2020) [21]	Predict tumour grade	139	-	139	pNET (pancreas)	MRI(T1, T2)	G1 vs G2/3	Pathology	-	A ML model with both radiomics and non-radiomics features has the best performance to predict tumour grade	47%
Canellas (2018) [22]	1) Predict tumour grade 2) predict progression	101	-	101	pNET (pancreas)	CT (PVP)	1) G1 vs G2/3 2) PFS	1) Pathology 2) > 20% increase in size OR new lesion	surgical resection	First-order features can predict tumour grade	0%
Chakraborty (2018) [23]	Predict tumour grade	55	-	55	pNET (pancreas)	CT (AP)	G1/2 vs G3	Pathology	-	A radiomics ML model is able to predict tumour grade	3%
Cheng (2019) [24]	Predict tumour grade	37	-	37	pNET (pancreas)	CT (nce, ce [phase not specified])	G1 vs G2/3	Pathology	-	First-order features are able to predict tumour grade	0%
Choi (2018) [25]	Predict tumour grade	66	-	66	pNET (pancreas)	CT (AP and PVP)	G1 vs G2/3	Pathology	-	Several radiomics features from both AP and PVP on 3D and 2D can predict tumour grade	0%
D'Onofrio (2019) [26]	Predict tumour grade	100	-	100	pNET (pancreas)	CT (AP+PVP)	G1 vs G2 vs G3	Pathology	-	First-order features can predict tumour grade	0%
De Robertis (2018) [27]	Predict tumour grade	42	-	42	pNET (pancreas)	MRI (T1, T1FS, T2, T2FS, late phase, post contrast, ADC)	G1 vs G2/3	Pathology	-	Several first-order features can predict tumour grade	0%
		106	-	106	pNET (pancreas)	MRI (T1-ce)	G1 vs G2 vs G3	Pathology	-		*

Table 1 (continued)

Authors	Aim	# Patients			Primary tumour (ROI)	Modality (sequence/phase)	Outcome	Reference standard	Intervention	Main results/conclusion	RQS
		GEP-NET	Other	Total							
Gao (2019) [28]	Predict tumour grade									Deep learning (CNN) can predict tumour grade	
Gu (2019) [29]	Predict tumour grade	138	-	138	pNET (pancreas)	CT (AP and PVP)	G1 vs G2/3	Pathology	-	A LR model with both radiomics and non-radiomics features has the best performance to predict tumour grade	53%
Guo (2019) [30]	Predict tumour grade	37	-	37	pNET (pancreas)	CT (AP)	G1/2 vs G3 (incl NEC)	Pathology	-	First-order features can predict pNET tumour grade	0%
Guo (2019) [31]	Predict tumour grade	77	-	77	pNET (pancreas)	MRI (T2, DWI)	G1 vs G2 vs G3	Pathology	-	A LR model is able to predict tumour grade	6%
Han (2021) [32]	Differentiate pNET from cysteadenoma	54	66	120	pNET/pCysteadenoma (pancreas)	CT (PVP)	pNET vs pCysteadenoma	Pathology	-	Several radiomics ML models can differentiate pNETs from pancreatic cysteadomas	28%
He (2019) [33]	Differentiate pNET from PDAC	147	-	147	pNET/PDAC (pancreas)	CT (AP)	pNET vs PDAC	Pathology	-	A ML model with either radiomics only or a combination of radiomics and non-radiomics features has the best performance to discriminate pNET from PDAC, compared to a model with non-radiomics features only	47%
Li (2018) [34]	Differentiate pNET from PDAC	25	50	75	pNET/PDAC (pancreas)	CT (PVP)	Hypovascular pNET vs PDAC	Pathology	-	First-order features are able to differentiate PDAC from pNET	0%
Li (2020) [35]	Differentiate pNET from SPN	61	58	119	pNET/SPN (pancreas)	MRI (T1, T2, ce[AP, PVP,DP], DWI, ADC)	pNET vs SPN	Pathology	-	A radiomics ML is better able to distinguish pNET from SPN, compared to radiological features only	22%
Liang (2019) [36]	Predict tumour grade	137	-	137	pNET (pancreas)	CT (AP)	G1 vs G2/3	Pathology	-	A radiomics nomogram combined with non-radiomics features has the best performance to predict pNET tumour grade	47%
Liang (2020) [59]	Predict tumour grade	61	-	61	Rectal NET (rectum)	CT (AP, PVP)	G1 vs G2/3/NEC	Pathology	-	Several first-order features can predict tumour grade	0%
Lin (2019) [37]	Differentiate pNET from accesory spleen	21	13	34	pNET/accessory Spleen (pancreas)	CT (AP)	pNET vs spleen	Pathology	-	First-order features can differentiate pNET from accessory spleen	0%
Liu (2021) [38]	Predict tumour grade	123	-	123	pNET (pancreas)	ceCT (AP) + MRI (T2, AP, PP, PVP)	G1 vs G2/3	Pathology	-	A ML model with both radiomics and non-radiomics features based on both CT or MRI can predict tumour grade	44%
Luo (2020) [39]	1) Predict tumour grade 2) predict OS	110	-	110	pNET (pancreas)	CT (AP, PVP)	1) G1/2 vs G3 2) Death	Pathology	-	A DL model can predict tumour grade, yet not significantly better than a traditional ML radiomics model	*
Mapelli (2020) [40]	1) predict tumour grade 2) predict RFS	61	-	61	pNET (pancreas)	[^{68}Ga]Ga- DOTATOC PET/CT [^{18}F]FDG PET/CT	1) G1 vs G2/3 2) recurrence 3) Angioinvasion	1) Pathology 2) NS	Surgical resection	None of the first-order features are able to predict tumour grade or RFS	0%

Table 1 (continued)

Authors	Aim	# Patients			Primary tumour (ROI)	Modality (sequence/phase)	Outcome	Reference standard	Intervention	Main results/conclusion	RQS
		GEP-NET	Other	Total							
							4) #lymphnodes				
Martini (2020) [57]	1) Predict tumour grade 2) Differentiate pNET from npNET 3) predict TTP 4) predict OS	23	25	48	pNET (livermetastases)	CT (AP, PVP)	1) pNET G1vsG2 2) pNET vs npNET 3) pNET Recurrence 4) pNET OS	1-2) Pathology 3) RECIST1.0 4) pNET Death	surgical resection	First-order features are able to predict TTP and OS	0%
Ohki (2020) [41]	Predict tumour grade	32	-	32	pNET (pancreas)	CT (AP, PVP) MRI (ADC)	G1 vs G2/3	Pathology	-	First-order features can predict pNET tumour grade	8%
Onner (2020) [58]	Predict response	22	-	22	pNET (All visible lesions)	[^{68}Ga]Ga-DOTATATE PET/CT	Response	Pathology	PRRT	First-order features can predict response to PRRT	0%
Pereira (2015) [42]	Predict tumour grade	22	-	22	pNET (pancreas)	MRI (ADC)	G1 vs G2 vs G3	Pathology	-	First-order features can predict pNET tumour grade	0%
Pulvirenti (2021) [43]	Predict tumour grade	150	-	150	pNET (pancreas)	CT (AP)	G1vsG2vsG3	Pathology	-	A LR radiomics model with both radiomics and non-radiomics features has the best performance to predict tumour grade	28%
Reinert (2020) [44]	Differentiate pNET from PDAC	42	53	95	pNET/PDAC (pancreas)	CT (PVP)	1) pNET vs PDAC 2) pNET G1vsG2/3	Pathology	-	A LR radiomics model can dfferentiate PDAC from PNET	6%
Shi (2020) [45]	Differentiate pNET from SPT	31	35	66	pNET/SPT (pancreas)	MRI (T2), (DWI: Dapp, Kapp)	pNET vs SPT	Pathology	-	A ML model with a combination of radiomics and non-radiomics features has the best performance to differentiate pNET from SPT, compared to subjective MRI diagnosis	50%
Shindo (2016) [46]	Differentiate pNET from PDAC	11	53	64	pNET/PDAC (pancreas)	MRI (ADC)	pNET vs PDAC	Pathology	-	First-order features can differentiate pNET from PDAC	0%
Song (2021) [47]	Predict recurrence	74	-	74	pNET (pancreas)	CT (AP, PVP)	Recurrence	Imaging	surgical resection	A DL model can predict recurrence with a better performance than traditional ML radiomics model	39%
Song (2021) [48]	Differentiate pNET from SPN	22	57	77	pNET (pancreas)	MRI (T1, AP, PVP, delayed phase)	pNET vs SPN	Pathology	-	A ML model based on MRI (AP) has the best performance in differentiating pNETs from SPNs compared to other MRI phases.	39%
van der Pol (2019) [49]	Differentiate pNET from RCC metastases	43	17	60	pNET/RCC (pancreas)	CT (PVP)	pNET vs RCC metastases	Pathology	-	First-order features can differentiate pNET from RCC pancreatic mets	0%
		18	32	40			pNET vs PDAC	Pathology	-		6%

Table 1 (continued)

Authors	Aim	# Patients			Primary tumour (ROI)	Modality (sequence/phase)	Outcome	Reference standard	Intervention	Main results/conclusion	RQS
		GEP-NET	Other	Total							
Wang (2019) [50]	Differentiate pNET from PDAC				pNET/PDAC (pancreas)	MRI (DWI IVIM: Dfast, Dslow, f)				A LR radiomics model can differentiate pNET from PDAC	
Wang (2020) [51]	Differentiate pNET from PDAC	21	63	84	pNET/PDAC (pancreas)	CT (Pancreatic, PVP)	pNET vs PDAC	Pathology	-	First order features can differentiate pNET from PDAC	28%
Werner (2019) [52]	1) Predict PFS 2) predict OS	31	-	31	pNET (All visible lesions)	[^{68}Ga]Ga- DOTATOC [^{68}Ga]Ga- DOTATATE PET/CT	1) PFS 2) OS	1) RECIST1.1 2) Death	PRRT	First order features can predict OS	0%
Xu (2019) [53]	Predict tumour grade	137	-	137	pNET (pancreas)	CT (PVP)	G1 vs G2/3	Pathology	-	A radiomics signature is able to predict tumour grade, but no significant difference between 2D or 3D based radiomics signature	28%
Yu (2020) [54]	Differentiate non--hypovascular pNET from PDAC	40	80	120	pNET/PDAC (pancreas)	CT (AP and PVP)	pNET vs PDAC	Pathology	-	A LR radiomics model based on the PVP is significantly better than AP or AP+PVP features and simple radiological features in differentiating pNET from PDAC	19%
Zhang (2021) [55]	Predict tumour grade	82	-	82	pNET (pancreas)	CT (AP)	G1 vs G2 vs G3	Pathology	-	Different ML models can predict tumour grade	31%
Zhao (2020) [56]	Predict tumour grade	99	-	99	pNET (pancreas)	CT (not specified)	G1 vs G2	Pathology	-	A radiomics ML model can predict tumour grade	44%

AP arterial phase, *ce* contrast-enhanced, *CT* computed tomography, *DL* deep learning, *DWI* diffusion-weighted imaging, *gNET* gastric NET, *G1/2/3* grade, *LR* logistic regression, *ML* machine learning, *MRI* magnetic resonance imaging, *NEC* neuroendocrine carcinoma, *NET* neuroendocrine tumour, *OS* overall survival, *pNET* pancreatic neuroendocrine tumour, *PDAC* pancreatic adenocarcinoma, *PET* positron emission tomography, *PFS* progression-free survival, *PVP* portal venous phase, *PRRT* peptide receptor radionuclide therapy, *RECIST 1.1* Response Evaluation Criteria in Solid Tumours, *SPN* solitary pseudopapillary neoplasms, *SPT* solid pseudopapillary tumour, *TTP* time to progression

*These are the studies using deep learning methods and are excluded from the RQS as described in the main text

Table 2 Details about radiomics analysis

Authors	Outcome	Segmentation method	#Observers	Software	Pre-processing method	# Features		Analyses	Validation
						Extracted	Selected		
Azoulay (2020) [15]	G3 vs NEC	Manual (2D)	1	TexRAD	n/a	72	72	MwU test, ROC AUC	No
Beleu (2020) [16]	G1/2 vs G3	Manual (2D)	?	LIFEx (5.10)	n/a	5	5	ANOVA/Kruskal Wallis, ROC curves, binary LR	No
Benedetti (2021) [17]	G1 vs G2/3	Manual (3D)	1	CGITA	Voxel normalisation and rebinning (64)	69		ROC AUC, MwU	No
Bevilacqua (2021) [18]	G1 vs G2	Manual (3D, 40% SUVmax)	1	Matlab	n/a	60		LDA, ROC/AUC	Unseen*
Bian (2020) [19]	G1 vs G2	Manual (3D)	2	PyRadiomics	n/a	1029		Kruskal-Wallis, X2, ROC curves, MLR	No
Bian (2020) [20]	G1 vs G2/3	Manual (3D)	2	PyRadiomics	Intensity normalisation, isotropic resampling	1409		Kruskal-Wallis, X2, ROC curves, LR	No
Bian (2020) [21]	G1 vs G2/3	Manual (3D)	3	PyRadiomics	Bias field correction, resampling, intensity standardization	2126		Rank sum test, X2, ROC AUC	Unseen
Canellas (2018) [22]	1) G1 vs G2/3 2) PFS	Manual (2D)	1	TexRAD	n/a	36		LR, ROC AUC	No
Chakraborty (2018) [23]	G1/2 vs G3	Manual (?)	?	?	Intensity normalisation	162		Naïve Bayes, RF, ROC	10-fold cross
Cheng (2019) [24]	G1 vs G2/3	Manual (2D)	1	TexRAD	Automatic motion correction	36		MwU	No
Choi (2018) [25]	G1 vs G2/3	Manual (2D+3D)	2	Medical imaging solution for segmentation and texture analyis	no	16		MwU/t-test, LR, ROC	No
D'Onofrio (2019) [26]	G1 vs G2 vs G3	Manual (3D)	2	Mazda	Reconstructed images (?)	5		Wilcox Mann-Whitman, ROC	No
De Robertis (2018) [27]	G1 vs G2/3	Manual (3D)	?	TensorFlow	n/a	14		MwU, ROC	No
Gao (2019) [28]	G1 vs G2 vs G3	Manual (3D), box	2	PyRadiomics	Isotropic resampling	-	-	CNN, GAN	External
Gu (2019) [29]	G1 vs G2/3	Manual (3D)	2	Omni-Kinetics	n/a	853		MLR, RF, ROC	External
Guo (2019) [30]	G1/2 vs G3 (incl NEC)	Manual (5 slices)	2	Inhouse, Matlab	n/a	5		Kruskal-Wallis, ANOVA, Spearman, ROC	No
Guo (2019) [31]	G1 vs G2 vs G3	Manual (3D)	2	LIFEx	n/a	68		Fisher's exact, LR, ROC	No

Table 2 (continued)

Authors	Outcome	Segmentation method	#Observers	Software	Pre-processing method	# Features		Analyses	Validation
						Extracted	Selected		
Han (2021) [32]	pNET vs pCysteadeno-ma	Semi-automatic (3D)	1	PyRadiomics	n/a	40		ML (LDA, SVM, RF, AdaBoost, KNN, GaussianNB, GBDT, LR, DT) with AUC	Unseen
He (2019) [33]	pNET vs PDAC	Manual (3D)	2	?	n/a	647	7	LASSO, SVM, RF, ROC	Unseen
Li (2018) [34]	hypovascular pNET vs PDAC	Manual (3D)	1	Mazda	Intensity normalisation	10		T-test/MwU, ROC AUC	No
Li (2020) [35]	pNET vs SPN	Manual (2D)	2	Fire Voxel	n/a	300		RDA, PCA, linear and non-linear discriminant analysis, MwU, ROC	Unseen
Liang (2019) [36]	G1 vs G2/3	Manual (3D)	1	Inhouse, Matlab	Intensity normalisation, isotropic resampling	467		LR	External
Liang (2020) [59]	G1 vs G2/3/NEC	Manual (3D)	2	Inhouse, Matlab	n/a	10		ROC AUC	No
Lin (2019) [37]	pNET vs spleen	Manual (3D, 5 slices)	2	PyRadiomics	Bias field correction, resampling, intensity standardization	12		MwU, ROC	No
Liu (2021) [38]	G1 vs G2/3	Manual (3D, multiple slices stacked)	1	Keras2.1.1	Voxel rescaling	1209		T-test/Kruskal Wallis (univariate), LDA, ROC/AUC	Unseen
Luo (2020) [39]	1) G1/2 vs G3 2) Death	Manual (3D rectangle) (DL) manual (3D) (ML)	?	CGITA	Rebinning (64)	> 8000		DL) CNN, 8-fold cross validation ML) SVM, LR, RF	External
Mapelli (2020) [40]	1) G1 vs G2/3 2) recurence 3) Angioinvasion 4) #lymphnodes	Automatic (3D) 40% SUVmax	?	TexRAD	n/a	7	7	LR	No
Martini (2020) [57]	1) pNET G1vsG2 2) pNET vs npNET 3) pNET Recurrence 4) pNET OS	Manual (3D)	2	Syngo	Intensity normalisation	36		MwU, Pearson, Cox LR	No
Ohki (2020) [41]	G1 vs G2/3	Manual (3D)	2	LIFEx (5.10)	n/a	50		ROC AUC	No
Onner (2020) [58]	Response	Semi-automatic (3D)	2 (consen-sus)	Image J	n/a	2		MwU, ROC	No
Pereira (2015) [42]	G1 vs G2 vs G3	Semi-automatic (3D)	1	?	Intensity normalisation	10		ANOVA, ROC	No
Pulvirenti (2021) [43]	G1 vs G2 vs G3	Manual (?)	1	Pyradiomics	Intensity normalisation	256		LR, SVM	Unseen

Table 2 (continued)

Authors	Outcome	Segmentation method	#Observers	Software	Pre-processing method	# Features		Analyses	Validation
						Extracted	Selected		
Reinert (2020) [44]	1) pNET vs PDAC 2) pNET G1 vs G2/3	Manual (2D)	2	Matlab	n/a	92		LR, ROC	No
Shi (2020) [45]	pNET vs SPT	Manual (3D)	2	Open-source (avllieres)	n/a	65		t-test, LR, ROC	Unseen
Shindo (2016) [46]	pNET vs PDAC	Manual (2D)	2	Synapse vincent	n/a	4		MwU, ROC	No
Song (2021) [47]	Recurrence	Manual (3D)	2	Inhouse	Image normalisation, isotropic resampling	143		SVM + 10-foldcrossvalidation, ROC, Kaplan-Meier	External
Song (2021) [48]	pNET vs SPN	Manual (3D)	2	PyRadiomics	Voxel resampling, grey-level discretion, image intensity normalisation.	396		mLR, ROC/AUC	Unseen
van der Pol (2019) [49]	pNET vs RCC metastases	Manual (2D)	2 (in 17%)	Image J	n/a	3		MLR, ROC	No
Wang (2019) [50]	pNET vs PDAC	Manual (2D), largest area	2	Inhouse (artificial intelligence Kit)	?	5		LR, ROC	No
Wang (2020) [51]	pNET vs PDAC	Manual (3D, 3 slices)	1	Image J	n/a	68		MwU/t-test, multivariable LR, ROC	Unseen
Werner (2019) [52]	1) PFS 2) OS	Manual (3D)	1	Interview Fusion Workstation	n/a	12		ROC, kaplan-Meier, Cox regression	No
Xu (2019) [53]	G1 vs G2/3	Manual (2D & 3D)	1	Inhouse, Matlab	Intensity normalisation, isotropic resampling	58		MwU, ROC, LR	Unseen
Yu (2020) [54]	pNET vs PDAC	Manual (3D)	1	Analysis-Kit	n/a	385		LR, ROC	No
Zhang (2021) [55]	G1 vs G2 vs G3	Manual (3D)	2	LIFEx	400-bit grey scale	40		ML (LDA, SVM, RF, AdaBoost, KNN, GaussianNB, LR, GBDT, DT)	Unseen
Zhao (2020) [56]	G1 vs G2	Manual (3D)	2	Inhouse, Matlab	Voxel rescaling	585		SVM, RBF, ROC, Kaplan-Meier, Cox regression	Unseen

2D delineation of 1 slice or only a part of the tumour volume, *3D* delineation of whole tumour volume, *AUC* area under the curve, *CNN* convolutional neural network, *DL* deep learning, *DT* decision tree, *G1/2/3* grade, *GAN* generative adversarial network, *GBDT* gradient boosting decision tree, *LR* logistic regression, *ML* machine learning, *MwU* Mann-Whitney U, *KNN* K-nearest neighbour, *OS* overall survival, *pNET* pancreatic neuroendocrine tumour, *PDAC* pancreatic adenocarcinoma, *PFS* progression-free survival, *RBF* radial basis function, *RF* random forest, *ROC* receiver operating curve, *SVM* support vector machine, X^2 chi^2 test

*Validation in the same institute on an unseen dataset; usually the dataset is split into training and validation.

2). Forty-three studies focused on patients with pancreas (p)NETs [14–55], two studies on GEP-NETs [56, 57], and one study on rectum NETs [58]. Thirty-one studies analysed CT [14–16, 18, 21–25, 28, 29, 31–33, 35–38, 40, 42, 43, 46, 48, 50, 52–56, 58], thirteen MRI [19, 20, 26, 27, 30, 34, 37, 40, 41, 44, 45, 47, 49], four [^{68}Ga]Ga-DOTATOC or [^{68}Ga]Ga-DOTATATE PET/CT [17, 39, 51, 57], and one [^{18}F]FDG PET/CT [39]. Differentiation of GEP-NETs from other lesions was investigated by 13 studies [14, 31, 32, 34, 36, 43–45, 47–50, 53, 59], while 25 explored GEP-NET grade [15–25, 27–30, 35, 37–40, 42, 52, 54, 55, 58], and 6 response to treatment or long-term outcome [21, 39, 46, 51, 56, 57]. The median number of included GEP-NET patients was 61 (range 11–157). A median of 58 (range 2–2126) radiomics features were extracted, excluding two DL studies [27, 38]. More than half (n = 27) of the studies adjusted for multiple testing or used feature reduction. Forty-one (91%) studies [14–30, 32–38, 40, 42–56, 58] used manual segmentation. Thirty (67%) studies delineated the entire tumour volume, yet it was not specified in 2 studies [22, 42]. Thirteen studies [14–16, 23, 25, 26, 34, 36, 40, 41, 45, 56, 57] analysed individual radiomics features (univariate), while thirty-two [17–22, 24, 27–33, 35, 37–39, 42–44, 46–55, 58, 60] developed models with multiple features. Seventeen studies performed multivariable analysis with non-radiomics features [18–21, 28, 29, 32, 34, 35, 37, 42, 44, 46, 47, 50, 53, 58].

Quality assessment

The included studies reached a median score of 2.0 points (range 4–19, Table 1 and S1), excluding two DL studies. The median total RQS was 2 (IQR 0–14), with a corresponding percentage of 5.6% (IQR 0–38.9%), due to a lack of prospective design, validation and open-access data. Noteworthy is that nineteen studies had a score of 0%. Most studies had a well-documented image protocol, included biological correlates, discussed potential clinical utility and compared results to a gold standard. Only one study employed a phantom to test the robustness of features [51]. Neither study had a prospective design, assessed cost-effectiveness or presented open-source scans or code. Only 11% of the studies externally validated their results, whereas 53% did not use any validation. Results of the subgroup analyses did not show a significant difference in neither aim nor imaging modality (Table 3). The RQS% distribution across the years is shown in Fig. 3.

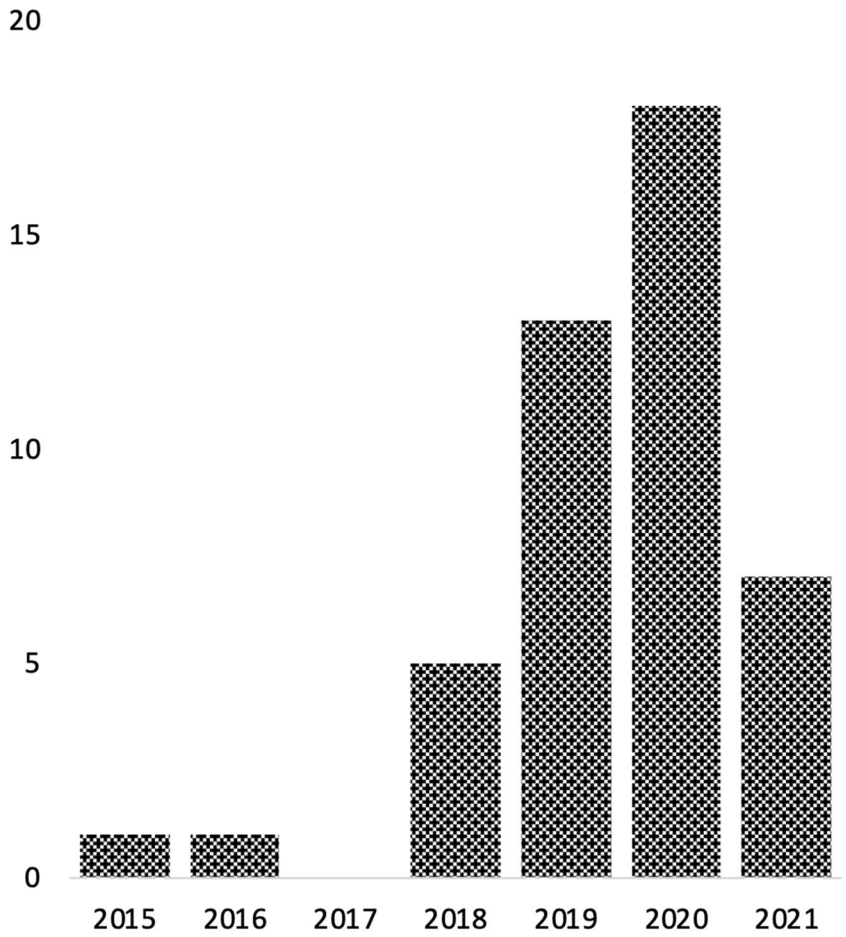

Fig. 2 Included radiomics studies in GEP-NETs, sorted by number of publications per year (until August 2021)

Table 3 Subgroup analysis according to the study aim and imaging modality

Group	Studies (n)	RQS	RQS-percentage	*p* value
Study aim				0.81
Differentiate	15	2.0 (0.0–10.0)	5.6 (0.0–27.8)	
Grade	25	1.0 (0.0–16.0)	5.6 (0.0–44.4)	
Response and long-term outcome	5	0.0 (0.0–7.0)	0.0 (0.0–19.4)	
Modality				0.46
CT	29	2.0 (0.0–12.5)	5.6 (0.0–34.7)	
MRI	12	5.0 (0.0–16.8)	15.3 (0.0–46.5)	
PET	4	0.0 (0.0–9.8)	0 (0.0–27.1)	

Values are expressed as number or median (interquartile range)

RQS radiomics quality score

Predicting tumour grade

Computed tomography

Four studies constructed models to predict tumour grade with CT (G1/2 vs G3, including poorly differentiated NECs) on arterial phase (AP) with a good performance (AUC 0.74–0.96 [22, 29, 38] and accuracy of 69% [42]). Seven studies constructed models for the prediction of G2/3 [18, 21, 24, 28, 35, 37, 38, 52]. Three radiomics models were able to predict G2/3 pNETs with a good performance (AUC 0.86–0.96, RQS 28–44%) [18, 52, 55]. One study compared different methods of feature selection and ML and found that distance correlation (feature selection) combined with AdaBoost (ML) had the best performance (G1vsG2: AUC 0.82, G2vsG3: AUC 0.70 and G1vsG3: AUC 0.85) [54]. Another study used LR to classify grades in rectal NETs (AUC 0.93) [58]. Three studies combined radiomics with non-radiomics features (e.g. size, vascular invasion, tumour margin) to create a good predictive model for G2/3 (AUC 0.75–0.90) [28, 35, 37]. In general, combined models yielded higher performance than models based on radiomics features only [21, 28, 35, 37], yet one study reported that performance was not improved by adding conventional radiographic features [42]. DL and traditional ML approaches had a similar performance to distinguish pNET grades [38]. Several radiomics features were selected as predictive variables in the constructed models [15, 21, 24, 28, 29, 35]. Both 2D- and 3D-rendered radiomics features on the portal venous phase (PVP) were predictive for G2/3 [24, 52]. First-order features were frequently selected, yet with conflicting results. Three studies found a lower entropy in high-grade tumours on AP [15, 29] with higher uniformity in G3 tumours [15], while a higher entropy was reported in G2/3 on PVP [21]. The abovementioned studies had a low sample size and were of low quality (RQS < 10%). Studies with a high RQS developed promising models, and predominantly higher-order features were selected for the prediction of pNET tumour grade [18, 28, 35, 37, 54, 55]. Only the features skewness [18, 37], GLRLM_GLV [28, 35], GLSZM_GreyLevelVariance [28, 35], and SizeZoneNonUniformity [28, 37] (with different filters or transformation methods) were selected in > 1 models.

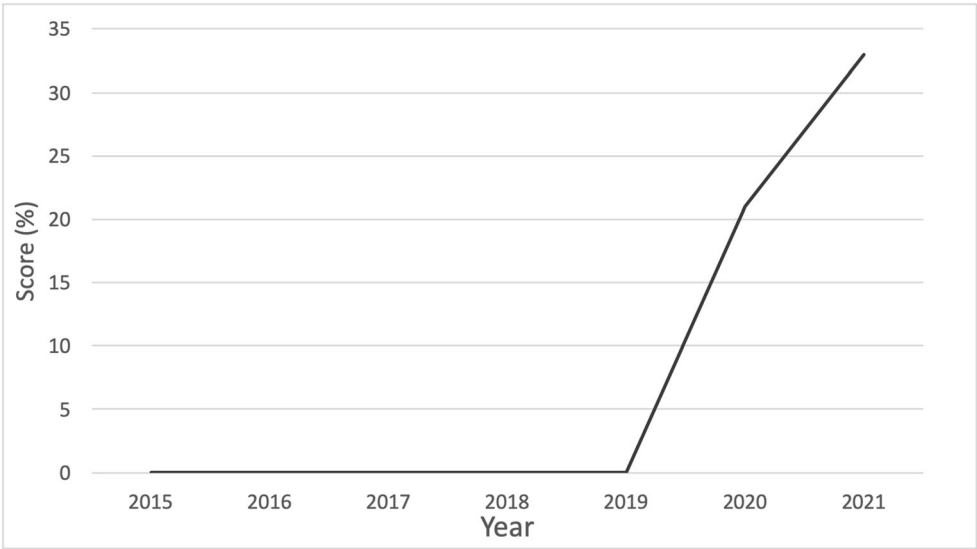

Fig. 3 Line plot of the RQS% scores of the included articles per year

Individual radiomics features were frequently tested in univariate analysis. A majority analysed first-order features, including mean [14, 23–25, 29, 40, 52, 56], entropy [14, 21, 23, 25, 40, 52, 56, 58], uniformity [15, 29], skewness [14, 23, 24, 52, 56], kurtosis [24, 25, 29, 40, 52, 56], sphericity [24] and asphericity [16]. Kurtosis was higher in G2/3 on AP [25, 29, 40, 56], yet conflicting results were reported for PVP [24, 40, 52, 56]. A higher skewness was reported in G2/3 tumours on different phases [23, 24, 52], while no differences between G1 and G2 pNETs were reported [14]. Conflicting results were reported for entropy [14, 21, 23, 25, 40, 52, 56] and mean [14, 24, 25, 29, 40, 52, 56].

Magnetic resonance imaging

Eight studies used MRI for grade prediction [19, 20, 26, 27, 30, 37, 40, 41], predominantly for G2/3. Several studies were able to G2/3 based on radiomics features only (AUC 0.74–0.91) [19, 20, 27]. Two higher-quality studies constructed a combined model with a good performance to predict G2/3 (AUC 0.71–0.88) [19, 37]. One lower-quality study was able to predict G3 tumours with a high performance (AUC0.99) [30]. Several features were selected as independent predictors in the constructed models. Kurtosis was frequently reported as a predictor of tumour grade [19, 20, 26, 41], yet with conflicting results. Squareroot_glszm_SmallAreaLowGreyLevelEmphasis was selected in two studies as a relevant predictor of G2/3 [19, 37]. Various features were selected in a single study only (Table S2).

Predominantly first-order features were analysed in univariate analysis. Kurtosis was not predictive of tumour grade on ADC [41,42] or T2w imaging [26]. A higher skewness was reported between G3 and G1, but no differences were found between G1 and G2 on ADC [41]. Conflicting results were reported for mean [40, 41], entropy [26, 30, 40], uniformity [26] and percentiles [26, 41] on different sequences (T1/T2/ADC/DWI).

PET/CT

One study used PET/CT-based radiomics and reported a high performance to predict pNET grade (G1vsG2, AUC 0.90–0.92) [17].

Distinguishing NET from other tumours

Computed tomography

Five studies constructed a model to differentiate pNET from pancreatic ductal adenocarcinoma (PDAC) on CT [32, 33, 43, 50, 53]. Overall, combined models achieved a better performance than radiomics-only models (AUC 0.83–0.88 vs. AUC0.79–0.87, respectively) [32, 50]. A LR model based on PVP-based features had a higher performance to distinguish pNETs from PDAC compared to an AP-based model (AUC0.93 vs AUC0.86, respectively) [53]. A comparable performance for LASSO, SVM and RF models was reported in a high-quality study[32]. A lower-quality study was able to distinguish pNET from PDAC (AUC 0.89) [33]. First-order radiomics features [33, 43, 50] and GLRLM features [32, 53] were most frequently selected as predictive features. Specific combinations of feature selection and ML methods (distance correlation, Xgboost+RF) had the best performance to differentiate pNET from pancreatic cystadenoma on PVP (AUC0.997 and AUC0.989) [31].

Predominantly first-order and GLCM features were analysed in univariate analysis [33, 43, 50, 53, 59]. pNETs had a higher mean [33], higher median [43, 50], higher minimum [43, 50], higher percentiles [33, 43, 50], higher entropy [50], and lower skewness [33], compared to PDAC. Similarly, a higher mean and lower skewness were reported when pNETs were compared with non-pNETs [56]. Other features were selected in a single study only (Table S2). Two studies compared pNETs to tumours from another origin [36, 48] and found that several first-order radiomics features were able to distinguish pNET from renal cell carcinoma (RCC) [48] or SPN [36].

Magnetic resonance imaging

MRI was used for differentiation of pNET in 5 studies [34, 44, 45, 47, 49]. Three studies developed a model for the differentiation of pNET from other pancreatic tumours [34, 44, 49]. A LR-model was able to differentiate pNETs from PDAC on Intravoxel Incoherent Motion (IVIM), yet was of low quality (AUC 0.93) [49]. Three studies constructed a radiomics model to differentiate pNET from solitary pseudo-papillary neoplasms (SPN) of the pancreas [34, 44, 47]. Significantly better performance of a radiomics model was reported, compared to subjective evaluation by a radiologist (accuracy 86–92% vs 65–78%, respectively) [34, 44]. An AP-based model had the best performance to differentiate pNET from SPN, compared to T1, PVP and delayed phases (AUC 0.91 vs. AUC 0.77–0.85, respectively) [47]. In the constructed models, first-order features were most predictive on T1W [34] and T2W [44], GLCM features for DWI [34], and GLRLM and NGTDM features were most predictive on the apparent diffusion and apparent kurtosis [44].

Two studies analysed individual radiomics features between pNET and PDAC [45, 49] on different sequences. Predominantly, first-order features were studied. A lower entropy [45], lower skewness [45] and lower kurtosis [45] were reported in pNET compared to PDAC on ADC. Likewise, a lower entropy was found in pNET on IVIM [49].

Response and long-term outcome measures

In 6 studies, response to treatment or long-term outcome was analysed [21, 39, 46, 51, 56, 57].

A higher skewness and kurtosis were found in non-responders to PRRT in univariate analysis, based on ^{68}Ga-SSA PET/CT [57]. A DL model based on CT-based radiomics features had a higher performance to predict recurrence in pNETs (AUC 0.77) compared to traditional ML methods, or DL including clinical features [46]. A higher entropy was reported in patients with better PFS and OS [51] on ^{68}Ga-SSA PET/CT in patients treated with PRRT, while on contrast-enhanced CT, a lower entropy was found in patients with a better PFS [21, 56]. No features were predictive of recurrence-free survival (RFS) based on [^{68}Ga]Ga-DOTATOC PET/CT [39]. Regarding OS, a higher skewness on CT was reported in patients with a better OS [56]. Different radiomics features were predictive of vascular invasion on CT and [^{68}Ga]Ga-DOTATOC PET/CT (Table S2) [16, 39].

Discussion

Because radiomics in GEP-NETs is still in its early stages, methodology is less standardised and studies are more explorative compared to radiomics studies in other fields. A majority of the included studies investigated GEP-NET grade and differentiation of GEP-NETs from other neoplasms, while a minority analysed the response to treatment or long-term outcomes. Strikingly, 43 studies analysed pNETs, while only 2 studied other GEP-NETs, underlining the lack of evidence of radiomics studies for GEP-NETs other than in the pancreas. A majority of the included studies were of low quality (RQS < 30%), mainly due to explorative and univariable analyses with a lack of (external) validation, feature reduction, calibration and/or bootstrapping. Radiomics reviews in other tumours show a similar trend [11, 12, 61], with regard to the lack of (external) validation, calibration and comprehensive models including non-radiomics features in their models. The predictive power of radiomics may be overestimated in these lower-quality studies, as there is a risk of false positives if no feature selection is used or only univariable analyses are performed, because of the high feature-to-patient ratio. In addition, the risk of overfitting is high if no validation is employed. Studies with a higher quality (RQS > 30%) were all published quite recently (2019 or later) [17–19, 28, 32, 35, 37, 44, 46, 47, 50, 54, 55, 62]. In these studies, the best predictive models were constructed for the prediction of tumour grade, in which combined models achieved the best performance. A minority showed promising results for radiomics in the differentiation of pNETs from other tumours, yet these results still need to be reproduced in larger cohort studies to ensure reliability. Thus, for the aforementioned outcomes, some promising models were developed, yet none performed a prospective validation.

Overall, prediction of tumour grade was mainly studied in pNETs with promising results on different modalities [17–22, 24, 27–30, 35, 37, 38, 52, 54]. First-order features were predominantly studied with conflicting results, but in general, an increased heterogeneity (higher entropy, kurtosis, max intensity and lower energy) was associated with higher-grade tumours. This is in line with biological studies that report that heterogeneous tumours, in general, have a worse outcome [63, 64]. It is noteworthy that the included studies used different versions of the World Health Organization (WHO) classification of NETs of i.e. 2010 [65], 2017 [66] and 2019 [67], which could have biased the results, since G3 well-differentiated tumours were only distinguished from poorly differentiated NECs in WHO 2017 and later.

Regarding the differentiation of GEP-NETs from other neoplasms, the best performance was achieved when radiomics features were combined with clinical features. Some studies even showed that a radiomics model performed better than a radiologist in differentiating pNETs from other lesions [34, 44]. The strongest evidence is available for distinguishing pNETs from PDACs on CT [32, 50, 53]. First-order statistics [33, 43, 45, 49, 50] and GLRLM features [32, 44, 53] were the best predictors in multivariable models.

From a clinical point of view, prediction of response or long-term outcome would be more relevant to explore, whereas only 13% studied these outcomes [21, 39, 46, 51, 56, 57] with conflicting results. Entropy was frequently reported as a predictor and might be worth further exploring. Nevertheless, evidence for the prediction of these outcomes with radiomics is still weak. An accurate prediction tool would allow for a timely adjustment of treatment strategy in GEP-NETs, and therefore, there is a need for further research exploring a more robust method for these outcomes.

An important prerequisite to enable high-quality reproducible radiomics research is to standardise the methodology as much as possible when a new model is constructed, according to guidelines suggested by the Image Biomarker Standardisation Initiative (IBSI) [68, 69]. One of the limitations that is harder to overcome in GEP-NET radiomics studies lies in the fact that GEP-NETs are quite rare and it is difficult to achieve a large homogeneous sample and validate studies. The lack of large samples could explain why the number of DL studies is low in this group of patients (n = 2), since this is a prerequisite for DL. Another limitation is the fact that in 98% of the included studies positive results were highlighted, while only one study did not find any predictive features for the studied outcomes. This is in line with a previous radiomics review (including 553 studies on GEP-NETs and other tumours) [70] and is likely to be a reflection of publication bias. The publication of negative findings in the field of radiomics is equally important to that of positive results to understand the directions for meaningful research that will bring the field to the next level. Future studies should focus on performing multicentre studies to develop integrated models, to be able to externally validate their models and explore DL.

In conclusion, the majority of radiomics studies in GEP-NETs is of low quality, which warrants new studies with a better methodology. Even though radiomics for GEP-NETs is still in its infancy, some robust and promising models have been developed.

However, these models predominantly focus on the identification of tumour grade or differentiation of GEP-NETs from other tumours and only few have externally validated their results. Finally, the quality of the studies that used radiomics for the prediction of response or long-term outcome, clinically more relevant endpoints, was quite low and more robust analyses are warranted before any definitive conclusions can be drawn.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-022-08996-w.

Funding The authors state that this work has not received any funding.

Declarations

Guarantor The scientific guarantor of this publication is R. G. H. Beets-Tan.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was not required for this study because of the nature of the study (i.e. systematic review of the literature).

Ethical approval Institutional Review Board approval was not required because of the nature of the study (i.e. systematic review of the literature).

Methodology

- Retrospective
- Systematic review
- performed at one institution

References

1. Marusyk A, Almendro V, Polyak K (2012) Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer 12:323
2. Marusyk A, Polyak K (2010) Tumor heterogeneity: causes and consequences. Biochim Biophys Acta 1805:105–117
3. Keck KJ, Choi A, Maxwell JE et al (2017) Increased grade in neuroendocrine tumor metastases negatively impacts survival. Ann Surg Oncol 24:2206–2212
4. Walter D, Harter PN, Battke F et al (2018) Genetic heterogeneity of primary lesion and metastasis in small intestine neuroendocrine tumors. Sci Rep 8:3811
5. Yang Z, Tang LH, Klimstra DS (2011) Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification. Am J Surg Pathol 35:853–860
6. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
7. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248
8. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762
9. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006
10. Limkin EJ, Sun R, Dercle L et al (2017) Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Ann Oncol 28:1191–1206
11. Ponsiglione A, Stanzione A, Cuocolo R et al (2022) Cardiac CT and MRI radiomics: systematic review of the literature and radiomics quality score assessment. Eur Radiol 32:2629–2638
12. Spadarella G, Ugga L, Calareso G, Villa R, D'Aniello S, Cuocolo R (2022) The impact of radiomics for human papillomavirus status prediction in oropharyngeal cancer: systematic review and radiomics quality score assessment. Neuroradiology. https://doi.org/10.1007/s00234-022-02959-0
13. McInnes MDF, Moher D, Thombs BD et al (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA Statement. JAMA 319:388–396
14. Azoulay A, Cros J, Vullierme MP et al (2020) Morphological imaging and CT histogram analysis to differentiate pancreatic neuroendocrine tumor grade 3 from neuroendocrine carcinoma. Diagn Interv Imaging 101:821–830
15. Beleù A, Rizzo G, De Robertis R et al (2020) Liver tumor burden in pancreatic neuroendocrine tumors: CT features and texture analysis in the prediction of tumor grade and (18)F-FDG uptake. Cancers (Basel) 12
16. Benedetti G, Mori M, Panzeri MM et al (2021) CT-derived radiomic features to discriminate histologic characteristics of pancreatic neuroendocrine tumors. Radiol Med 126:745–760
17. Bevilacqua A, Calabro D, Malavasi S et al (2021) A [68Ga]Ga-DOTANOC PET/CT Radiomic model for non-invasive prediction of tumour grade in pancreatic neuroendocrine tumours. Diagnostics (Basel) 11.5: 870.
18. Bian Y, Jiang H, Ma C et al (2020) CT-Based radiomics score for distinguishing between grade 1 and grade 2 nonfunctioning pancreatic neuroendocrine tumors. AJR Am J Roentgenol 215:852–863
19. Bian Y, Li J, Cao K et al (2020) Magnetic resonance imaging radiomic analysis can preoperatively predict G1 and G2/3 grades in patients with NF-pNETs. Abdom Radiol (NY) 46:667–680
20. Bian Y, Zhao Z, Jiang H et al (2020) Noncontrast radiomics approach for predicting grades of nonfunctional pancreatic neuroendocrine tumors. J Magn Reson Imaging 52:1124–1136
21. Canellas R, Burk KS, Parakh A, Sahani DV (2018) Prediction of pancreatic neuroendocrine tumor grade based on CT features and texture analysis. AJR Am J Roentgenol 210:341–346
22. Chakraborty J, Pulvirenti A, Yamashita R et al (2018) Quantitative CT analysis for the preoperative prediction of pathologic grade in pancreatic neuroendocrine tumors. SPIE, Department of Surgery, United States Department of Radiology, United States Department of Epidemiology and Biostatistics, United States Department of Pathology, United States Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, United States
23. Cheng SH, Jin ZY, Xue HD (2019) Evaluation of the histologic grade of pancreatic neuroendocrine tumors using CT texture analysis and perfusion parameters. Int J Clin Exp Med 12:771–779
24. Choi TW, Kim JH, Yu MH, Park SJ, Han JK (2018) Pancreatic neuroendocrine tumor: prediction of the tumor grade using CT

findings and computerized texture analysis. Acta Radiol 59:383-392
25. D'Onofrio M, Ciaravino V, Cardobi N et al (2019) CT enhancement and 3D texture analysis of pancreatic neuroendocrine neoplasms. Sci Rep 9:2176
26. De Robertis R, Maris B, Cardobi N et al (2018) Can histogram analysis of MR images predict aggressiveness in pancreatic neuroendocrine tumors? Eur Radiol 28:2582–2591
27. Gao X, Wang X (2019) Deep learning for World Health Organization grades of pancreatic neuroendocrine tumors on contrast-enhanced magnetic resonance images: a preliminary study. Int J Comput Assist Radiol Surg 14:1981–1991
28. Gu D, Hu Y, Ding H et al (2019) CT radiomics may predict the grade of pancreatic neuroendocrine tumors: a multicenter study. Eur Radiol 29:6880–6890
29. Guo C, Zhuge X, Wang Z et al (2019) Textural analysis on contrast-enhanced CT in pancreatic neuroendocrine neoplasms: association with WHO grade. Abdom Radiol (NY) 44:576–585
30. Guo C-G, Ren S, Chen X et al (2019) Pancreatic neuroendocrine tumor: prediction of the tumor grade using magnetic resonance imaging findings and texture analysis with 3-T magnetic resonance. Cancer Manag Res 11:1933–1944
31. Han X, Yang J, Luo J et al (2021) Application of CT-based radiomics in discriminating pancreatic cystadenomas from pancreatic neuroendocrine tumors using machine learning methods. Front Oncol 11:606677
32. He M, Liu Z, Lin Y et al (2019) Differentiation of atypical non-functional pancreatic neuroendocrine tumor and pancreatic ductal adenocarcinoma using CT based radiomics. Eur J Radiol 117:102–111
33. Li J, Lu J, Liang P et al (2018) Differentiation of atypical pancreatic neuroendocrine tumors from pancreatic ductal adenocarcinomas: Using whole-tumor CT texture analysis as quantitative biomarkers. Cancer Med 7:4924–4931
34. Li X, Zhu H, Qian X, Chen N, Lin X (2020) MRI texture analysis for differentiating nonfunctional pancreatic neuroendocrine neoplasms from solid pseudopapillary neoplasms of the pancreas. Acad Radiol 27:815–823
35. Liang W, Yang P, Huang R et al (2019) A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. Clin Cancer Res: 25(2): 584-594
36. Lin X, Xu L, Wu A, Guo C, Chen X, Wang Z (2019) Differentiation of intrapancreatic accessory spleen from small hypervascular neuroendocrine tumor of the pancreas: textural analysis on contrast-enhanced computed tomography. Acta Radiol 60: 553–560
37. Liu C, Bian Y, Meng Y et al (2021) Preoperative Prediction of G1 and G2/3 Grades in patients with nonfunctional pancreatic neuroendocrine tumors using multimodality imaging. Acad Radiol. https://doi.org/10.1016/j.acra.2021.05.017
38. Luo Y, Chen X, Chen J et al (2020) Preoperative prediction of pancreatic neuroendocrine neoplasms grading based on enhanced computed tomography imaging: validation of deep learning with a convolutional neural network. Neuroendocrinology 110:338–350
39. Mapelli P, Partelli S, Salgarello M et al (2020) Dual tracer 68Ga-DOTATOC and 18F-FDG PET/computed tomography radiomics in pancreatic neuroendocrine neoplasms: an endearing tool for preoperative risk assessment. Nucl Med Commun 41:896–905.
40. Ohki K, Igarashi T, Ashida H et al (2020) Usefulness of texture analysis for grading pancreatic neuroendocrine tumors on contrast-enhanced computed tomography and apparent diffusion coefficient maps. Jpn J Rad. https://doi.org/10.1007/s11604-020-01038-9
41. Pereira JAS, Rosado E, Bali M, Metens T, Chao S-L (2015) Pancreatic neuroendocrine tumors: correlation between histogram analysis of apparent diffusion coefficient maps and tumor grade. Abdom Imaging 40:3122–3128
42. Pulvirenti A, Yamashita R, Chakraborty J et al (2021) Quantitative computed tomography image analysis to predict pancreatic neuroendocrine tumor grade. JCO Clin Cancer Inform 5:679–694
43. Reinert CP, Baumgartner K, Hepp T, Bitzer M, Horger M (2020) Complementary role of computed tomography texture analysis for differentiation of pancreatic ductal adenocarcinoma from pancreatic neuroendocrine tumors in the portal-venous enhancement phase. Abdom Radiol (NY) 45:750–758
44. Shi YJ, Zhu HT, Liu YL et al (2020) Radiomics analysis based on diffusion kurtosis imaging and T2 weighted imaging for differentiation of pancreatic neuroendocrine tumors from solid pseudopapillary tumors. Front Oncol 10:1624
45. Shindo T, Fukukura Y, Umanodan T et al (2016) Histogram analysis of apparent diffusion coefficient in differentiating pancreatic adenocarcinoma and neuroendocrine tumor. Medicine (Baltimore) 95:e2574
46. Song C, Wang M, Luo Y et al (2021) Predicting the recurrence risk of pancreatic neuroendocrine neoplasms after radical resection using deep learning radiomics with preoperative computed tomography images. Ann Transl Med 9:833
47. Song T, Zhang Q-W, Duan S-F et al (2021) MRI-based radiomics approach for differentiation of hypovascular non-functional pancreatic neuroendocrine tumors and solid pseudopapillary neoplasms of the pancreas. BMC Med Imaging 21:36
48. van der Pol CB, Lee S, Tsai S et al (2019) Differentiation of pancreatic neuroendocrine tumors from pancreas renal cell carcinoma metastases on CT using qualitative and quantitative features. Abdom Radiol (NY) 44:992–999
49. Wang YW, Zhang XH, Wang BT et al (2019) Value of texture analysis of intravoxel incoherent motion parameters in differential diagnosis of pancreatic neuroendocrine tumor and pancreatic adenocarcinoma. Chin Med Sci J 34:1-9
50. Wang Z, Chen X, Wang J, Cui W, Ren S, Wang Z (2020) Differentiating hypovascular pancreatic neuroendocrine tumors from pancreatic ductal adenocarcinoma based on CT texture analysis. Acta Radiol 61:595–604
51. Werner RA, Ilhan H, Lehner S et al (2019) Pre-therapy somatostatin receptor-based heterogeneity predicts overall survival in pancreatic neuroendocrine tumor patients undergoing peptide receptor radionuclide therapy. Mol Imaging Biol 21:582–590
52. Xu L, Yang P, Yen EA et al (2019) A multi-organ cancer study of the classification performance using 2D and 3D image features in radiomics analysis. Phys Med Biol 64:215009
53. Yu H, Huang Z, Li M et al (2020) Differential diagnosis of nonhypervascular pancreatic neuroendocrine neoplasms from pancreatic ductal adenocarcinomas, based on computed tomography radiological features and texture analysis. Acad Radiol 27:332–341
54. Zhang T, Zhang Y, Liu X et al (2020) Application of radiomics analysis based on CT combined with machine learning in diagnostic of pancreatic neuroendocrine tumors patient's pathological grades. Front Oncol 10:521831
55. Zhao Z, Bian Y, Jiang H et al (2020) CT-radiomic approach to predict G1/2 nonfunctional pancreatic neuroendocrine tumor. Acad Radiol 27:e272–e281. https://doi.org/10.1016/j.acra.2020.01.002
56. Martini I, Polici M, Zerunian M et al (2020) CT texture analysis of liver metastases in PNETs versus NPNETs: correlation with histopathological findings. Eur J Radiol 124:108812
57. Onner H, Abdulrezzak U, Tutus A (2020) Could the skewness and kurtosis texture parameters of lesions obtained from pretreatment Ga-68 DOTA-TATE PET/CT images predict receptor radionuclide therapy response in patients with gastroenteropancreatic neuroendocrine tumors? Nuclear Med Commun 41:1034–1039
58. Liang P, Xu C, Tan F et al (2020) Prediction of the World Health Organization Grade of rectal neuroendocrine tumors based on CT histogram analysis. Cancer Med 10.2: 595-604

59. Guo C, Zhuge X, Wang Q et al (2018) The differentiation of pancreatic neuroendocrine carcinoma from pancreatic ductal adenocarcinoma: the values of CT imaging features and texture analysis. Cancer Imaging 18:37
60. Wang R, Liu H, Liang P, Zhao H, Li L, Gao J (2021) Radiomics analysis of CT imaging for differentiating gastric neuroendocrine carcinomas from gastric adenocarcinomas. Eur J Radiol 138: 109662
61. Chetan MR, Gleeson FV (2021) Radiomics in predicting treatment response in non-small-cell lung cancer: current status, challenges and future perspectives. Eur Rad 31:1049–1058
62. Bian Y, Zhao Z, Jiang H et al (2020) Noncontrast radiomics approach for predicting grades of nonfunctional pancreatic neuroendocrine tumors. J Magn Reson Imaging 52:1124-1136 : https://doi.org/10.1002/jmri.27176
63. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C (2015) Translational implications of tumor heterogeneity. Clin Cancer Res: 21:1258–1266
64. Morris LG, Riaz N, Desrichard A et al (2016) Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. Oncotarget 7:10051–10063
65. Fléjou JF (2011) WHO Classification of digestive tumors: the fourth edition. Ann Pathol 31:S27–S31
66. Lloyd RV, Osamura RY, Kloppel G, Rosai J (2017) WHO classification of tumours of endocrine organs. 4th Edition, Volume 10
67. Nagtegaal ID, Odze RD, Klimstra D et al (2020) The 2019 WHO classification of tumours of the digestive system. Histopathology 76:182-188
68. Zwanenburg A, Leger S, Vallières M, Löck S (2016) Image biomarker standardisation initiative. arXiv:1612.07003
69. Zwanenburg A, Vallières M, Abdalah MA et al (2020) The image biomarker standardization initiative: standardised quantitative radiomics for high-throughput image-based phenotyping. Radiology 295:328–338
70. Song J, Yin Y, Wang H, Chang Z, Liu Z, Cui L (2020) A review of original articles published in the emerging field of radiomics. Eur J Radiol 127:108991