MDPI

Article

Artificial Intelligence-Based Segmentation of Residual Tumor in Histopathology of Pancreatic Cancer after Neoadjuvant Treatment

Boris V. Janssen [1,2], Rutger Theijse [1,2], Stijn van Roessel [1], Rik de Ruiter [3], Antonie Berkel [3], Joost Huiskens [3], Olivier R. Busch [1], Johanna W. Wilmink [4], Geert Kazemier [5], Pieter Valkema [2], Arantza Farina [2], Joanne Verheij [2,†], Onno J. de Boer [2,†] and Marc G. Besselink [1,*,†]

[1] Department of Surgery, Amsterdam UMC, Cancer Center Amsterdam, University of Amsterdam, 1081 HV Amsterdam, The Netherlands; b.v.janssen@amsterdamumc.nl (B.V.J.); r.t.theijse@amsterdamumc.nl (R.T.); s.vanroessel@amsterdamumc.nl (S.v.R.); o.r.busch@amsterdamumc.nl (O.R.B.)
[2] Department of Pathology, Amsterdam UMC, Cancer Center Amsterdam, University of Amsterdam, 1081 HV Amsterdam, The Netherlands; p.valkema@amsterdamumc.nl (P.V.); a.farina@amsterdamumc.nl (A.F.); j.verheij@amsterdamumc.nl (J.V.); o.j.deboer@amsterdamumc.nl (O.J.d.B.)
[3] SAS Institute Besloten Vennootschap, 1272 PC Huizen, The Netherlands; Rik.deRuiter@sas.com (R.d.R.); Antonie.berkel@sas.com (A.B.); Joost.Huiskens@SAS.com (J.H.)
[4] Department of Medical Oncology, Amsterdam UMC, Cancer Center Amsterdam, University of Amsterdam, 1081 HV Amsterdam, The Netherlands; j.w.wilmink@amsterdamumc.nl
[5] Department of Surgery, Amsterdam UMC, Cancer Center Amsterdam, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands; g.kazemier@amsterdamumc.nl
[*] Correspondence: m.g.besselink@amsterdamumc.nl; Tel.: +31-20-444-4444
[†] These authors share senior authorship.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Simple Summary: The use of neoadjuvant therapy (NAT) in patients with pancreatic ductal adenocarcinoma (PDAC) is increasing. Objective quantification of the histopathological response to NAT may be used to guide adjuvant treatment and compare the efficacy of neoadjuvant regimens. However, current tumor response scoring (TRS) systems suffer from interobserver variability, originating from subjective definitions, the sometimes challenging histology, and response heterogeneity throughout the tumor bed. This study investigates if artificial intelligence-based segmentation of residual tumor burden in histopathology of PDAC after NAT may offer a more objective and reproducible TRS solution.

Abstract: Background: Histologic examination of resected pancreatic cancer after neoadjuvant therapy (NAT) is used to assess the effect of NAT and may guide the choice for adjuvant treatment. However, evaluating residual tumor burden in pancreatic cancer is challenging given tumor response heterogeneity and challenging histomorphology. Artificial intelligence techniques may offer a more reproducible approach. Methods: From 64 patients, one H&E-stained slide of resected pancreatic cancer after NAT was digitized. Three separate classes were manually outlined in each slide (i.e., tumor, normal ducts, and remaining epithelium). Corresponding segmentation masks and patches were generated and distributed over training, validation, and test sets. Modified U-nets with varying encoders were trained, and F1 scores were obtained to express segmentation accuracy. Results: The highest mean segmentation accuracy was obtained using modified U-nets with a DenseNet161 encoder. Tumor tissue was segmented with a high mean F1 score of 0.86, while the overall multiclass average F1 score was 0.82. Conclusions: This study shows that artificial intelligence-based assessment of residual tumor burden is feasible given the promising obtained F1 scores for tumor segmentation. This model could be developed into a tool for the objective evaluation of the response to NAT and may potentially guide the choice for adjuvant treatment.

Keywords: pancreatic cancer; histopathology; tumor response scoring; neoadjuvant therapy; artificial intelligence; machine learning

1. Background

Neoadjuvant therapy (NAT) is increasingly used for patients with locally advanced and (borderline) resectable pancreatic ductal adenocarcinoma (PDAC). Recent clinical studies have shown that NAT affects overall survival, disease-free survival, and margin-negative resection rates positively [1–6]. Histologic examination of PDAC resection specimens following NAT offers the opportunity to evaluate treatment response using various tumor response scoring (TRS) systems [7,8]. TRS is believed to serve at least two important purposes [7]. First, it may be used in clinical trials to compare the efficacy of different neoadjuvant regimens in PDAC. Secondly, TRS may guide the choice for adjuvant therapy in the individual patient. It is imperative that TRS accurately correlates with the oncological outcome (i.e., overall survival) for both purposes.

Over the last few decades, researchers proposed several histopathological TRS systems for PDAC to evaluate NAT responses [8–10]. Still, during the 2019 Amsterdam international consensus meeting on histological assessment of tumor response of resected pancreatic cancer after NAT, the international study group of pancreatic pathologists (ISGPP) stated that most TRS systems suffer from flawed reasoning or lack objective definition criteria [7]. Most fundamentally, the interobserver agreement is highly insufficient for the most used systems or has not been widely determined, likely related to the subjective nature of the criteria used to define the different categories. Some TRS systems evaluate the ratio of vital tumor rests versus treatment-induced fibrosis. However, distinguishing between treatment-induced fibrosis and desmoplasia or pancreatitis-related fibrosis is at the very least challenging and may, in fact, be impossible with the naked eye. Because of these limitations, the ISGPP stated that a new TRS system should assess residual (viable) tumor burden instead of tumor regression, and objective criteria for the different categories are needed. Artificial intelligence (AI)-based techniques have the potential to fulfill these needs. Artificial intelligence models may be developed to automatically segment and quantify residual tumor burden in histological sections of neoadjuvantly treated and resected PDAC, potentially providing the basis for an objective TRS system.

In this study, we investigated if AI-based segmentation of residual tumor burden in histopathological slides after resection of PDAC after NAT is feasible and may offer a foundation for a more objective and reproducible solution for TRS. To this end, we report developing an AI-based segmentation tool in PDAC segmenting residual tumor burden in histopathological slides of patients following NAT to study treatment response.

2. Materials and Methods

2.1. Data Acquisition

We retrospectively collected histopathological hematoxylin and eosin (H&E)-stained slides of pancreatic cancer resection specimens of neoadjuvantly treated patients from the archive of the Department of Pathology at Amsterdam UMC in the Netherlands. Any amount or type of neoadjuvant chemo(radio)therapy was considered suitable to be included in this study. Per patient, one representative H&E slide of the tumor bed was selected by A.F. or J.V. and digitized using a Philips Intellisite Ultra-Fast Scanner (Philips, Best, The Netherlands). Whole-slide images (WSI) were converted to BigTiff format and downloaded from the image management system.

2.2. Data Handling

First, an expert pathologist (A.F. or J.V.) marked a representative region of interest (ROI) on the WSI using the 'ASAP' software package [11]. Within the boundaries of these ROI, histopathological structures were manually outlined at the pixel level using ASAP with coordinates stored as XML. The following structures were annotated: (1) normal ducts; (2) cancerous ducts; (3) in situ neoplasia; (4) islets of Langerhans; (5) acinic tissue; (6) atrophic metaplastic parenchyma; (7) fat; (8) vessels; (9) nerves; and (10) lymphocytic infiltrates. Detailed annotations were prepared in ASAP by B.J., R.T., and A.F., and when made by non-pathologists (B.J. and R.T.), they were evaluated and, if necessary, corrected

by expert pancreas pathologists. If a case was considered too difficult to annotate manually reliably, it was excluded from further analysis.

The ROIs were converted from BigTiff to PNG format, and the same area ground-truth annotation-based segmentation masks were generated as PNG files. Images and corresponding masks were generated at a resolution of 0.5 μm/px ('20×'). The original classifications were pooled into three new classes for further analysis: (1) normal ducts; (2) cancerous ducts and in situ neoplasia combined; and (3) the remaining non-tumorous epithelial tissue (NTET), consisting of islets of Langerhans, acinic tissue, and atrophic metaplastic parenchyma. The remaining classes (fat, vessels, nerves, and lymphocytic infiltrates) were ignored and considered as background elements. Using a sliding window approach, partly (50%) overlapping patches of 512 by 512 pixels were generated from the H&E and corresponding mask image. Patches and corresponding mask images were only included in the dataset if at least 10% of the patch's surface area was occupied by one of the segmentation classes. The maximum number of patches for each class was limited to 100 per case to limit class imbalance in the training data.

2.3. Machine Learning

A 'standard' U-net [12] and a selection of modified U-nets with different encoders, including ResNet158, EfficientNet-b1, -b4, and -b7, DenseNet161 and −201, all pre-trained on ImageNet, were used [13]. RGB intensity values of the H&E images were normalized during training, and data augmentation was performed by performing random rotations (90, 180, and 270 degrees) or horizontal or vertical flips. Binary cross-entropy was used as a loss function, combined with the ADAM optimizer, using a learning rate of 1×10^{-5} and weight decay of 0.1 every 10 epochs. Networks were trained for 30 epochs, and training was stopped if the validation error did not improve for 7 epochs.

After training, the test set predictions were made using a sliding window approach, followed by combining neighboring patches to a full ROI prediction. To avoid stitching artifacts in the reconstructed prediction, we generated partly overlapping patches. The weighted average of the segmentation probabilities was calculated and converted to either binary or RGB prediction mask images, essentially as described by Cui et al. [14]. The prediction accuracy of the test set was calculated for each class separately and expressed as an F1 (also known as Dice) score. Machine learning and data handling were performed with Python 3.6 and Pytorch 1.7 using one RTX3090 graphics processing unit with 24 GB of internal memory.

3. Results

3.1. Dataset

Histopathological H&E-stained slides of 65 pancreatic cancer resection specimens were collected. One specimen was not included because it was not feasible to segment it manually due to its growth pattern. Of the 64 remaining specimens, 43 patients were treated with FOLFIRINOX, 19 with gemcitabine-based chemoradiotherapy, and 2 with gemcitabine in combination with nab-paclitaxel. Table 1 details all the included cases. From the 64 specimens, 50 were used for training, 5 for validation, and 9 as a test set. Of all the generated patches, approximately 3000 contained normal ducts against 6000 with tumor tissue and 5000 with the remaining class, NTET. In total, 14,328 patches were used for training and 2244 for validation.

Table 1. Patient characteristics.

Characteristic	Number (*n*)	Percentage (%)
Tumor Location		
Head	54	84.4
Body	6	9.4
Tail	4	6.3
Neoadjuvant Therapy		
FOLFIRINOX × 8	20	31.3
FOLFIRINOX × 4	20	31.3
Gemcitabine × 3 + RTx × 1	19	29.7
Gem-nab-paclitaxel	2	3.1
FOLFIRINOX × 6	1	1.6
FOLFIRINOX × 2	1	1.6
FOLFIRINOX × 1	1	1.6

Legend: FOLFIRINOX = combined therapy of irinotecan, 5-fluorouracil, leucovorin, and oxaliplatin; RTx = radiotherapy.

3.2. AI-Based Histopathological Classification of Pancreatic Tissue

The accuracy of the obtained algorithms for segmenting the tumor, normal pancreatic ducts, and NTET in the H&E-stained histopathological sections is summarized in Table 2. The differences between the best-performing models were minor, and the best mean results were obtained when using the U-nets with a DenseNet161 encoder. The best results for tumor segmentation only were obtained using a U-net trained using a ResNet152 encoder. Notably, the modified U-nets with EfficientNet encoders performed relatively poorly; the mean F1 scores for the tumor class and NTET were the lowest of all investigated models. Moreover, these latter models appeared unable to segment normal ducts. The classic U-net performed better than the EfficientNet variants but was approximately 5% less accurate than the DenseNet or ResNet variants. In Figure 1, representative examples of three different cases of the AI-based segmentation results on previously unseen test samples are illustrated. In the figure, the H&E staining, the ground truth (as labeled by the expert pathologists), and the model's prediction (U-net with DenseNet161 encoder) are illustrated. From this figure, it can be appreciated that the major (and large) structures were recognized well, but clearly, some discrepancies between the ground truth labeling and the network's prediction were present.

Table 2. Obtained F1 scores using different U-net encoders.

Encoder	Tumor (F1, 95% CI)	Normal Ducts (F1, 95% CI)	NTET (F1, 95% CI)	Mean (F1)
DenseNet161	0.86 ± 0.09	0.74 ± 0.12	0.85 ± 0.07	0.82
DenseNet201	0.85 ± 0.09	0.77 ± 0.13	0.85 ± 0.08	0.82
EffecientNet-b1	0.78 ± 0.15	0	0.77 ± 0.13	0.51
EffecientNet-b4	0.77 ± 0.14	0	0.61 ± 0.73	0.46
EffecientNet-b7	0.81 ± 0.12	0	0.82 ± 0.12	0.54
ResNet152	0.88 ± 0.06	0.77 ± 0.14	0.73 ± 0.15	0.79
None ('standard' U-net)	0.83 ± 0.10	0.69 ± 0.23	0.83 ± 0.15	0.78

Legend: F1 = Dice score; 95% CI = 95% confidence interval; NTET = remaining non-tumor epithelium; Mean = unweighted average of F1 scores for the tumor, normal ducts, and NTET.
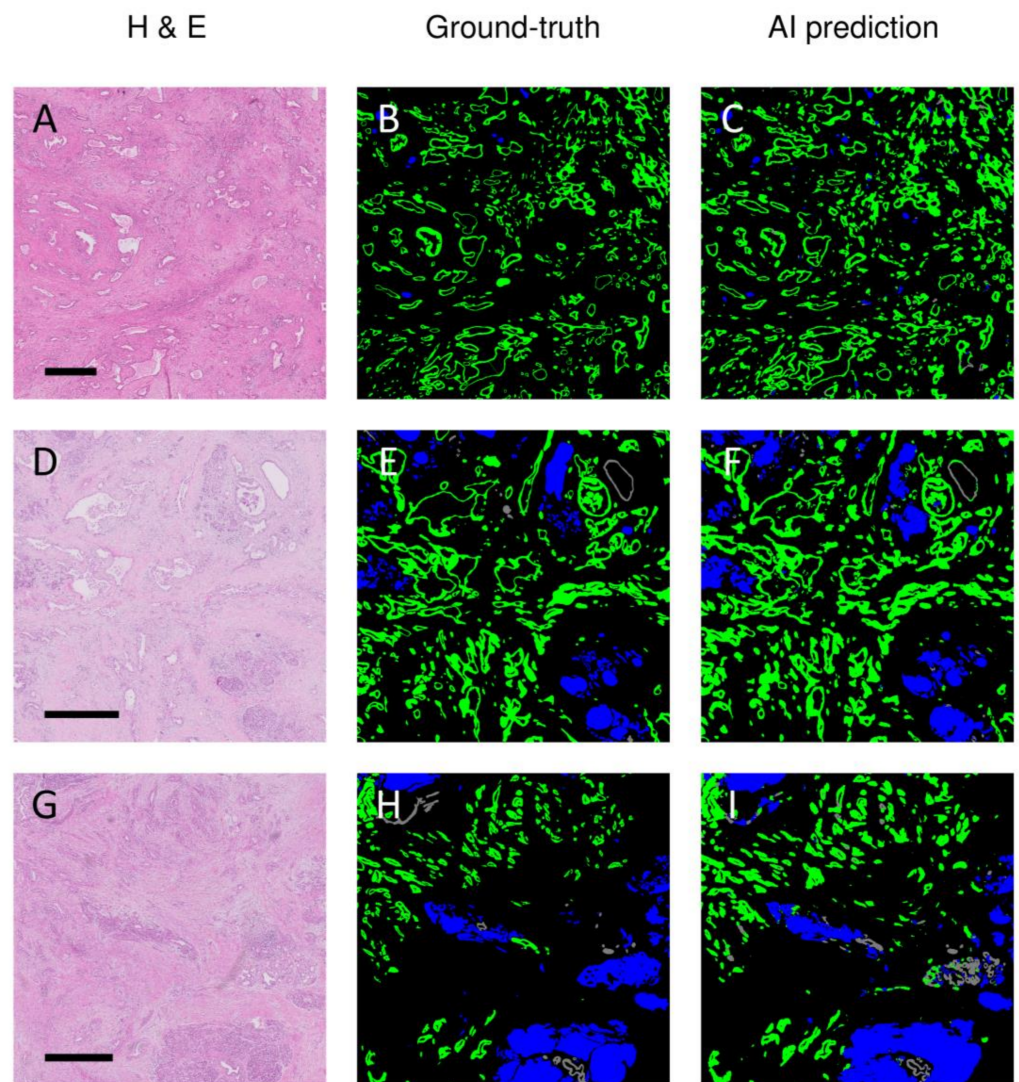
Figure 1. Representative examples of AI-based predictions versus pathologist-based ground-truth annotations of three different (previously unseen) patients with pancreatic ductal adenocarcinoma after neoadjuvant treatment. Legend: H&E = hematoxylin and eosin staining (**A,D,G**); Ground truth = annotations by the pathologist (**B,E,H**); AI prediction = prediction based on U-net with DenseNet161 encoder (**C,F,I**); green = cancerous ducts and in situ neoplasia; gray = normal ducts; blue = remaining non-tumorous epithelial tissue. The scale bar indicates a length of 1000 μm.

3.3. Discrepancies between Ground-Truth and AI-Based Predictions

To gain insight into the errors of the trained DenseNet161 network, the discrepancies between the ground-truth and AI predictions were carefully optically compared with the region of interest in the corresponding H&E section. Several kinds of mismatches were observed. First, the model frequently correctly identified structures that were not delineated during the annotation process, such as tumor buds (Figure 2A–C) or islets of Langerhans (Figure 2D–F). Second, the model incorrectly marked non-cancerous structures as cancerous. Figure 2G–I shows how the model incorrectly predicted ductal metaplasia as malignant ducts. On other occasions, the model erroneously marked vascular structures and folding artifacts as malignant ducts. Finally, the model also classified cancerous structures as non-cancerous, as illustrated in Figure 2J–L.
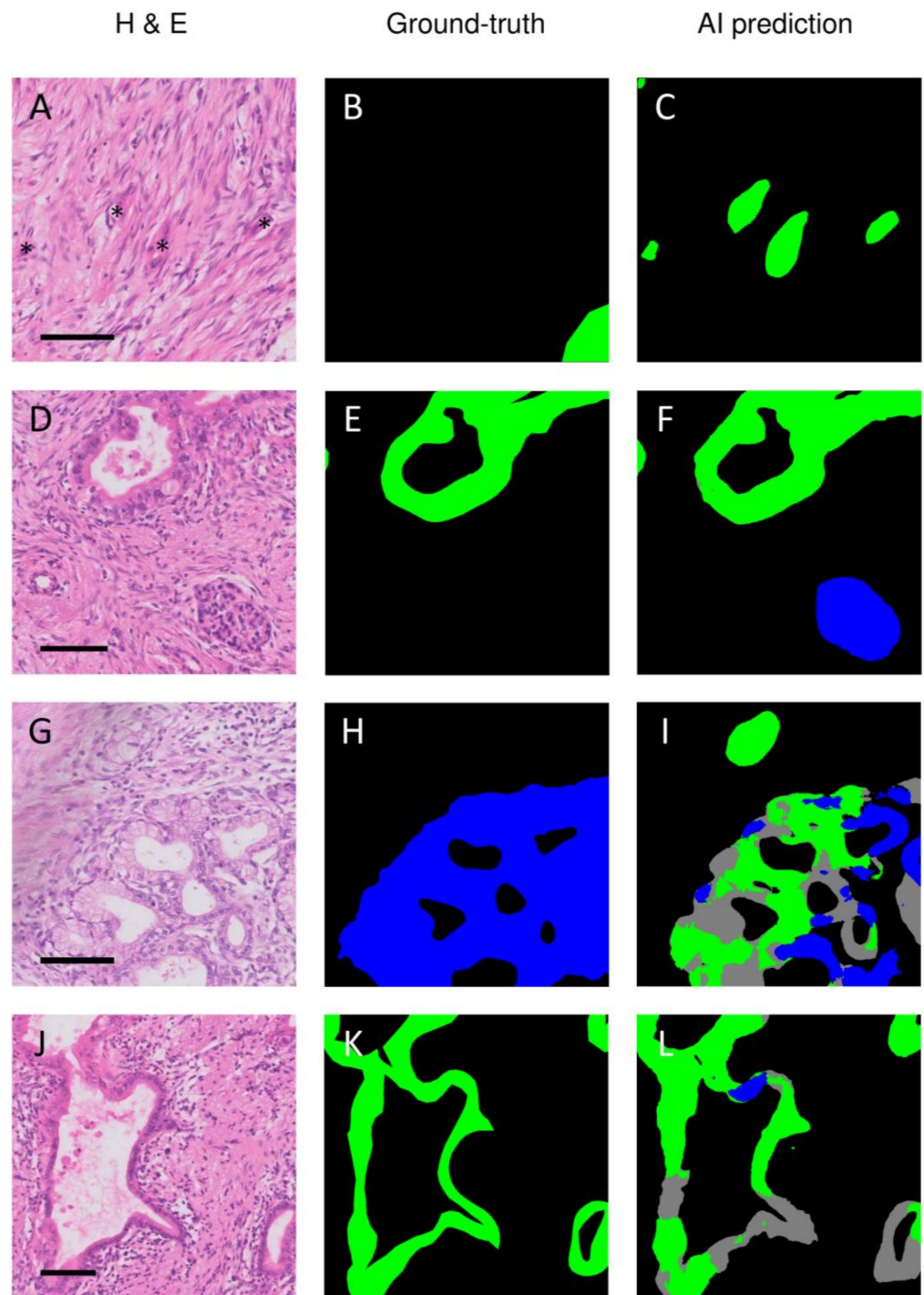
Figure 2. Discrepancies between the ground truth annotations and AI-based prediction of pancreatic ductal adenocarcinoma. Legend: The network frequently correctly recognized structures that were missed during annotating, such as tumor buds (**A–C**) or islets of Langerhans (**D–F**). The neural network also made classification errors, such as classifying atrophic metaplastic epithelium as either normal or tumorous ducts (**G–I**) or classifying tumor ducts as being normal (**J–L**). H&E = hematoxylin and eosin; Ground-truth = annotations by the pathologist; AI prediction = prediction by the artificial intelligence model; green = cancerous ducts and in situ neoplasia; gray = normal ducts; blue = remaining non-tumorous epithelial tissue. The asterisks in A indicate cancerous tissue that was missed during the annotation process. The scale bar indicates a length of 100 μm.

4. Discussion

This study shows that it is feasible to segment the residual tumor burden of PDAC after NAT using an AI-based approach. The highest mean segmentation accuracies were obtained with modified U-nets, especially when trained with a DenseNet161, −201, or ResNet152 encoder, all of which performed better than the standard U-net. Modified U-nets with an EfficientNet-b1, -b4, or -b7 encoder, on the other hand, performed considerably worse when compared with all other tested models, including the standard U-net.

To the best of our knowledge, this is the first report on an AI-based segmentation model to identify residual tumor burden after NAT in PDAC. There are AI-based segmentation models for colorectal, prostate, breast, liver, gastric, squamous, and basal cell cancers published in the peer-reviewed literature [15–37]. Clinical purposes for these segmentation tools vary, including detection and diagnosis, Gleason grading for prostate cancer, and prognostic or predictive feature extraction. In one study, an AI-based segmentation tool was used to measure the response to chemoradiotherapy in a cohort of patients with hepatocellular carcinoma [28]. Moreover, one study reported a segmentation tool that was developed in non-neoadjuvantly treated pancreatic cancer tissue [38]. This model reached a pixel-based precision and recall of 98.6% and 95.1%, respectively. At first sight, these scores appear to be better than our results in the present study. However, these data were obtained on consecutive slides of a single case. Therefore, their model is not directly comparable to ours. Their model was not developed to recognize the wide variety of heterogeneous presentations of PDAC after NAT, but rather to create a three-dimensional representation of histopathological micro-anatomy of one particular case. Therefore, model generalizability is likely to be poor. In terms of performance to similar studies in different cancer types, our model, with a mean F1 score of 0.86, compares favorably against the published performances, ranging from 0.353 to 0.9243 and a median F1 score of 0.835 [15–37]. When comparing our approach to quantifying residual tumor burden to current TRS systems in pancreatic cancer, several take a similar approach [9,10,39,40]. These systems aim to objectively quantify tumor burden and demonstrate favorable performance in predictive ability and reproducibility compared with some of the most-used TRS systems [9,10,40,41]. These studies indicate that objective quantification of residual tumor burden potentially forms a good basis for TRS. Still, for these systems, interobserver variation originating from human assessors remains an issue.

Although the highest unweighted mean (three class) accuracy was seen when we used modified U-nets with a DenseNet161 or −201 encoder, the difference with a ResNet152 encoder was slight. If we only considered the accuracy for identifying tumor tissue in the PDAC samples, the ResNet152 encoder, on average, marginally outperformed the DenseNet variants. When comparing mistakes made by these models, they each exhibited all error types, as shown in Figure 2. However, it should be noted that when we compared these three models (DenseNet161, DenseNet201, and ResNet152), there was not one model that performed best in all individual test cases. In some test cases, the DenseNet variants performed best, while in others, the ResNet variant achieved the highest scores. Given that no one model performed best in all individual test cases, we anticipate that the observed differences in the mean F1 scores were mainly the result of the composition of the relatively small test set, which may be more favorable to a single model. As such, we anticipate that performance differences may diminish when a larger and thus more representative test set is used.

On the other hand, we did consistently see that DenseNet- and ResNet-modified U-net models always performed better than the standard U-net. Interestingly, not all modified U-nets performed better than the standard U-net. All evaluated EfficientNet models performed considerably worse. The reason for this was not apparent. Still, maybe this had to do with the class imbalance in our dataset and that modified U-nets with EfficientNet encoders have more problems with underrepresented classes. Although we attempted to limit class imbalance during the creation of patches for training, it was a given that normal ducts in PDAC samples were relatively rare and not evenly distributed between individual cases.

Across all trained models, we always observed the lowest F1 scores for the normal duct class. This could be partially explained by the previously mentioned class imbalance in our training set. However, it is also important to note that even for the (experienced) human eye, it can be difficult to classify a specific duct as normal or abnormal solely based on morphology. Sometimes, aspects other than cytonuclear atypia are considered, like the architecture, the position within the tissue, and the surroundings. Therefore, it might not be surprising that an algorithm also has difficulty learning subtle characteristics of some malignancies.

The ground-truth annotations must be as accurate as possible to determine model performance. In retrospect, evaluation of the discrepancies between the manually annotated ground-truth and AI-based predictions revealed the manual segmentations to be suboptimal in several respects. In various instances, cancerous tissue was annotated inaccurately or not annotated at all. The latter was, for example, the case with tiny tumor buds, which were often correctly recognized by the algorithms but not manually annotated. In addition, other structures like the islets of Langerhans were occasionally not annotated. The lack of annotation of these structures was probably because annotations of small structures are very labor-intensive and tedious. On the other hand, the dignity of some structures was uncertain and therefore not annotated. It is difficult to predict the effect of these errors on the calculation of the F1 score or even on training the network. Structures missed in the ground-truth annotations but correctly predicted by the algorithms will therefore downgrade the F1 scores. Still, careful identification of the mismatches between the ground truth and the algorithm's prediction can help to improve the ground-truth annotations of our dataset, which ultimately will lead to improved performance of the algorithm, a technique known as active learning [42].

While the current results are promising, the present study has several limitations, and there is still plenty of room for improvement in segmentation performance and generalizability. For generalizability, it should be noted that the current dataset is relatively small and that all data came from the same laboratory. Additionally, all sections were digitized on only one type of scanner, and the same two pathologists essentially did all the annotation and classification. To develop generally applicable algorithms that accurately identify PDAC in histological samples, more data from different (international) laboratories with patients who received various neoadjuvant regimens must be included. Further improvement can be obtained if the annotation and classification of the pathologic structures are jointly performed by a larger group of expert pancreatic pathologists, and each tissue section is evaluated and classified by multiple experts. Indeed, tissue regions were frequently encountered that could not be classified as either normal or abnormal. Under these circumstances, the study could benefit from a consensus diagnosis from different experts. We should also note that we trained the current models on a relatively small dataset, and all cases were relatively well-differentiated, allowing manual delineation. It was practically impossible to manually annotate tumors with complex and reticular growth patterns accurately, and therefore, one of these cases was not included in the current dataset. Thus, the reported model performance might be optimistic for poorly differentiated tumors. Finally, to better understand factors influencing the prediction accuracies, such as tumor morphology, staining quality, and the age of the archived samples, future studies should ideally provide model performance data on each sample.

In addition to developing more generalizable and better-performing models using larger datasets, future research should optimize the data preparation and AI workflow. The quality of the ground truth could be improved by researching techniques such as antibody-supervised learning [43]. Immunohistochemistry-stained sections can typically provide better contrast compared to H&E-stained sections. This improved contrast can aid in identifying cells that could be mislabeled otherwise and may help to more easily create or automatically generate annotations that follow the contours of structures very precisely. This technique is especially relevant in cases with diffuse cancer growth and an abundance of solitary cancer cells or tumor buds. Moreover, though very labor-intensive, the quality of the ground truth could be further optimized using consensus-based segmentation of

the same cases by multiple pathologists. To expand on this in the view of (international) collaborative research, we could see additional benefits in online accessible segmentation platforms. Finally, with the rapid development of AI, researchers should stay vigilant for novel approaches to developing segmentation models. These approaches could involve ensemble learning techniques, weight optimization techniques, alternative data augmentations, and segmentation approaches based on active, semi-supervised, or unsupervised learning techniques.

Ultimately, a new AI-based TRS system quantifying residual tumor burden may address the issues of currently used TRS systems such as subjectivity, handling with response heterogeneity, and the inherent complexity of recognizing and quantifying diffuse cancerous growth and solitary cancer cells. We hypothesize that solving these problems will likely improve the clinical value of TRS. Next to providing a basis for an objective TRS system, we anticipate that this segmentation model could form the basis for tools that can extract relevant tissue features from segmentations on images to gain insight into cancers' molecular backgrounds. Insight into the molecular backgrounds of individual cancers could aid patient stratification by, for example, identifying novel biomarkers that correlate to the clinical outcome or by identifying molecular tumor subtypes with differential responses to neo(adjuvant) therapies [44–46].

5. Conclusions

We demonstrated that AI-based segmentation of residual tumor burden in pancreatic cancer after NAT is feasible and may form the basis for an objective TRS system. Further research is required, focusing on the development of extensive training databases and improving training data quality.

Author Contributions: Conceptualization, B.V.J., S.v.R., J.H., O.R.B., J.W.W., G.K., A.F., J.V., O.J.d.B. and M.G.B.; methodology, B.V.J., R.d.R., J.H., A.F, J.V., O.J.d.B. and M.G.B.; software, R.d.R., A.B., J.H., P.V., O.J.d.B.; validation, B.V.J., R.T., A.F., J.V., O.J.d.B. and M.G.B.; formal analysis, R.d.R., O.J.d.B.; investigation, B.V.J., S.v.R., A.F., J.V., O.J.d.B. and M.G.B.; data curation, B.V.J., R.T. and O.J.d.B.; writing—original draft preparation, B.V.J., R.T., S.v.R., A.F., J.V., O.J.d.B. and M.G.B.; writing—review and editing, B.V.J., R.T., S.v.R., R.d.R., A.B., J.H., O.R.B., J.W.W., G.K., P.V., A.F., J.V., O.J.d.B. and M.G.B.; visualization, B.V.J., A.F. and O.J.d.B.; supervision, S.v.R., J.H., A.F., J.V., O.J.d.B. and M.G.B.; project administration, B.V.J. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ryan, D.P.; Hong, T.S.; Bardeesy, N. Pancreatic adenocarcinoma. *N. Engl. J. Med.* **2014**, *371*, 1039–1049. [CrossRef]
2. Versteijne, E.; Suker, M.; Groothuis, K.; Akkermans-Vogelaar, J.M.; Besselink, M.G.; Bonsing, B.A.; Buijsen, J.; Busch, O.R.; Creemers, G.M.; van Dam, R.M.; et al. Preoperative Chemoradiotherapy Versus Immediate Surgery for Resectable and Borderline Resectable Pancreatic Cancer: Results of the Dutch Randomized Phase III PREOPANC Trial. *J. Clin. Oncol.* **2020**, *38*, 1763–1773. [CrossRef]
3. Versteijne, E.; Vogel, J.A.; Besselink, M.G.; Busch, O.R.C.; Wilmink, J.W.; Daams, J.G.; van Eijck, C.H.J.; Groot Koerkamp, B.; Rasch, C.R.N.; van Tienhoven, G. Meta-analysis comparing upfront surgery with neoadjuvant treatment in patients with resectable or borderline resectable pancreatic cancer. *Br. J. Surg.* **2018**, *105*, 946–958. [CrossRef] [PubMed]

4. Jang, J.Y.; Han, Y.; Lee, H.; Kim, S.W.; Kwon, W.; Lee, K.H.; Oh, D.Y.; Chie, E.K.; Lee, J.M.; Heo, J.S.; et al. Oncological Benefits of Neoadjuvant Chemoradiation With Gemcitabine Versus Upfront Surgery in Patients with Borderline Resectable Pancreatic Cancer: A Prospective, Randomized, Open-label, Multicenter Phase 2/3 Trial. *Ann. Surg.* **2018**, *268*, 215–222. [CrossRef] [PubMed]

5. Suker, M.; Beumer, B.R.; Sadot, E.; Marthey, L.; Faris, J.E.; Mellon, E.A.; El-Rayes, B.F.; Wang-Gillam, A.; Lacy, J.; Hosein, P.J.; et al. FOLFIRINOX for locally advanced pancreatic cancer: A systematic review and patient-level meta-analysis. *Lancet Oncol.* **2016**, *17*, 801–810. [CrossRef]

6. Van Eijck, C.H.J.; Versteijne, E.; Suker, M.; Groothuis, K.; Besselink, M.G.H.; Busch, O.R.C.; Bonsing, B.A.; Groot Koerkamp, B.; de Hingh, I.H.J.T.; Festen, S.; et al. Preoperative chemoradiotherapy to improve overall survival in pancreatic cancer: Long-term results of the multicenter randomized phase III PREOPANC trial. *J. Clin. Oncol.* **2021**, *39*, 4016. [CrossRef]

7. Janssen, B.V.; Tutucu, F.; van Roessel, S.; Adsay, V.; Basturk, O.; Campbell, F.; Doglioni, C.; Esposito, I.; Feakins, R.; Fukushima, N.; et al. Amsterdam International Consensus Meeting: Tumor response scoring in the pathology assessment of resected pancreatic cancer after neoadjuvant therapy. *Mod. Pathol.* **2021**, *34*, 4–12. [CrossRef]

8. van Roessel, S.; Janssen, B.V.; Soer, E.C.; Fariña Sarasqueta, A.; Verbeke, C.S.; Luchini, C.; Brosens, L.A.A.; Verheij, J.; Besselink, M.G. Scoring of tumour response after neoadjuvant therapy in resected pancreatic cancer: Systematic review. *Br. J. Surg.* **2021**, *108*, 119–127. [CrossRef]

9. Chou, A.; Ahadi, M.; Arena, J.; Sioson, L.; Sheen, A.; Fuchs, T.L.; Pavlakis, N.; Clarke, S.; Kneebone, A.; Hruby, G.; et al. A Critical Assessment of Postneoadjuvant Therapy Pancreatic Cancer Regression Grading Schemes With a Proposal for a Novel Approach. *Am. J. Surg. Pathol.* **2021**, *45*, 394–404. [CrossRef]

10. Neyaz, A.; Tabb, E.S.; Shih, A.; Zhao, Q.; Shroff, S.; Taylor, M.S.; Rickelt, S.; Wo, J.Y.; Fernandez-Del Castillo, C.; Qadan, M.; et al. Pancreatic ductal adenocarcinoma: Tumour regression grading following neoadjuvant FOLFIRINOX and radiation. *Histopathology* **2020**, *77*, 35–45. [CrossRef]

11. Litjens, G. Automated Slide Analysis Platform. Available online: https://computationalpathologygroup.github.io/ASAP/ (accessed on 5 June 2021).

12. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*; Springer: Cham, Switzerland, 2015; pp. 234–241.

13. Yakubovskiy, P. Segmentation Models Pytorch. Available online: https://github.com/qubvel/segmentation_models.pytorch (accessed on 14 May 2021).

14. Cui, Y.; Zhang, G.; Liu, Z.; Xiong, Z.; Hu, J. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. *Med. Biol. Eng. Comput.* **2019**, *57*, 2027–2043. [CrossRef] [PubMed]

15. Abdelsamea, M.M.; Grineviciute, R.B.; Besusparis, J.; Cham, S.; Pitiot, A.; Laurinavicius, A.; Ilyas, M. Tumour parcellation and quantification (TuPaQ): A tool for refining biomarker analysis through rapid and automated segmentation of tumour epithelium. *Histopathology* **2019**, *74*, 1045–1054. [CrossRef]

16. Amgad, M.; Sarkar, A.; Srinivas, C.; Redman, R.; Ratra, S.; Bechert, C.J.; Calhoun, B.C.; Mrazeck, K.; Kurkure, U.; Cooper, L.A.; et al. Joint Region and Nucleus Segmentation for Characterization of Tumor Infiltrating Lymphocytes in Breast Cancer. *Proc. SPIE Int. Soc. Opt. Eng.* **2019**, *10956*, 109560M. [CrossRef]

17. Bulten, W.; Bándi, P.; Hoven, J.; van de Loo, R.; Lotz, J.; Weiss, N.; van de Laak, J.; van Ginneken, B.; Hulsbergen-van de Kaa, C.; Litjens, G. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Sci. Rep.* **2019**, *9*, 864. [CrossRef]

18. Chen, C.M.; Huang, Y.S.; Fang, P.W.; Liang, C.W.; Chang, R.F. A computer-aided diagnosis system for differentiation and delineation of malignant regions on whole-slide prostate histopathology image using spatial statistics and multidimensional DenseNet. *Med. Phys.* **2020**, *47*, 1021–1033. [CrossRef]

19. Chen, Y.; Janowczyk, A.; Madabhushi, A. Quantitative Assessment of the Effects of Compression on Deep Learning in Digital Pathology Image Analysis. *JCO Clin. Cancer Inform.* **2020**, *4*, 221–233. [CrossRef]

20. Feng, R.; Liu, X.; Chen, J.; Chen, D.Z.; Gao, H.; Wu, J. A Deep Learning Approach for Colonoscopy Pathology WSI Analysis: Accurate Segmentation and Classification. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 3700–3708. [CrossRef] [PubMed]

21. Guo, Z.; Liu, H.; Ni, H.; Wang, X.; Su, M.; Guo, W.; Wang, K.; Jiang, T.; Qian, Y. A Fast and Refined Cancer Regions Segmentation Framework in Whole-slide Breast Pathological Images. *Sci. Rep.* **2019**, *9*, 882. [CrossRef] [PubMed]

22. Haj-Hassan, H.; Chaddad, A.; Harkouss, Y.; Desrosiers, C.; Toews, M.; Tanougast, C. Classifications of multispectral colorectal cancer tissues using convolution neural network. *J. Pathol. Inform.* **2017**, *8*, 1.

23. Ho, D.J.; Yarlagadda, D.V.; D'Alfonso, T.M.; Hanna, M.G.; Grabenstetter, A.; Ntiamoah, P.; Brogi, E.; Tan, L.K.; Fuchs, T.J. Deep multi-magnification networks for multi-class breast cancer image segmentation. *Comput. Med. Imaging Graph.* **2021**, *88*, 101866. [CrossRef]

24. Li, J.; Speier, W.; Ho, K.C.; Sarma, K.V.; Gertych, A.; Knudsen, B.S.; Arnold, C.W. An EM-based semi-supervised deep learning approach for semantic segmentation of histopathological images from radical prostatectomies. *Comput. Med. Imaging Graph.* **2018**, *69*, 125–133. [CrossRef] [PubMed]

25. Li, S.; Jiang, H.; Yao, Y.-d.; Pang, W.; Sun, Q.; Kuang, L. Structure convolutional extreme learning machine and case-based shape template for HCC nucleus segmentation. *Neurocomputing* **2018**, *312*, 9–26. [CrossRef]

26.   Mavuduru, A.; Halicek, M.; Shahedi, M.; Little, J.V.; Chen, A.Y.; Myers, L.L.; Fei, B. Using a 22-Layer U-Net to Perform Segmentation of Squamous Cell Carcinoma on Digitized Head and Neck Histological Images. *Proc. SPIE Int. Soc. Opt. Eng.* **2020**, *11320*, 113200C. [CrossRef]

27.   Qaiser, T.; Tsang, Y.-W.; Taniyama, D.; Sakamoto, N.; Nakane, K.; Epstein, D.; Rajpoot, N. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Med. Image Anal.* **2019**, *55*, 1–14. [CrossRef] [PubMed]

28.   Roy, M.; Kong, J.; Kashyap, S.; Pastore, V.P.; Wang, F.; Wong, K.C.; Mukherjee, V. Convolutional autoencoder based model HistoCAE for segmentation of viable tumor regions in liver whole-slide images. *Sci. Rep.* **2021**, *11*, 1–10. [CrossRef] [PubMed]

29.   Salvi, M.; Molinari, F.; Dogliani, N.; Bosco, M. Automatic discrimination of neoplastic epithelium and stromal response in breast carcinoma. *Comput. Biol. Med.* **2019**, *110*, 8–14. [CrossRef]

30.   Silva-Rodríguez, J.; Colomer, A.; Naranjo, V. WeGleNet: A weakly-supervised convolutional neural network for the semantic segmentation of Gleason grades in prostate histology images. *Comput. Med. Imaging Graph.* **2021**, *88*, 101846. [CrossRef]

31.   Sun, C.; Li, C.; Zhang, J.; Rahaman, M.M.; Ai, S.; Chen, H.; Kulwa, F.; Li, Y.; Li, X.; Jiang, T. Gastric histopathology image segmentation using a hierarchical conditional random field. *Biocybern. Biomed. Eng.* **2020**, *40*, 1535–1555. [CrossRef]

32.   van Rijthoven, M.; Balkenhol, M.; Siliņa, K.; van der Laak, J.; Ciompi, F. HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. *Med. Image Anal.* **2021**, *68*, 101890. [CrossRef] [PubMed]

33.   van Zon, M.C.M.; van der Waa, J.D.; Veta, M.; Krekels, G.A.M. Whole-slide margin control through deep learning in Mohs micrographic surgery for basal cell carcinoma. *Exp. Dermatol.* **2021**, *30*, 733–738. [CrossRef] [PubMed]

34.   Wang, X.; Fang, Y.; Yang, S.; Zhu, D.; Wang, M.; Zhang, J.; Tong, K.Y.; Han, X. A hybrid network for automatic hepatocellular carcinoma segmentation in H&E-stained whole slide images. *Med. Image Anal.* **2021**, *68*, 101914. [CrossRef]

35.   Xu, L.; Walker, B.; Liang, P.-I.; Tong, Y.; Xu, C.; Su, Y.C.; Karsan, A. Colorectal Cancer Detection Based on Deep Learning. *J. Pathol. Inform.* **2020**, *11*, 28. [CrossRef]

36.   Xu, Y.; Zhu, J.-Y.; Chang, E.I.C.; Lai, M.; Tu, Z. Weakly supervised histopathology cancer image segmentation and classification. *Med. Image Anal.* **2014**, *18*, 591–604. [CrossRef] [PubMed]

37.   Yang, Q.; Xu, Z.; Liao, C.; Cai, J.; Huang, Y.; Chen, H.; Tao, X.; Huang, Z.; Chen, J.; Dong, J.; et al. Epithelium segmentation and automated Gleason grading of prostate cancer via deep learning in label-free multiphoton microscopic images. *J. Biophotonics* **2020**, *13*, e201900203. [CrossRef] [PubMed]

38.   Kiemen, A.; Braxton, A.M.; Grahn, M.P.; Han, K.S.; Babu, J.M.; Reichel, R.; Amoa, F.; Hong, S.-M.; Cornish, T.C.; Thompson, E.D.; et al. In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution. *bioRxiv* **2020**. [CrossRef]

39.   Okubo, S.; Kojima, M.; Matsuda, Y.; Hioki, M.; Shimizu, Y.; Toyama, H.; Morinaga, S.; Gotohda, N.; Uesaka, K.; Ishii, G.; et al. Area of residual tumor (ART) can predict prognosis after post neoadjuvant therapy resection for pancreatic ductal adenocarcinoma. *Sci. Rep.* **2019**, *9*, 17145. [CrossRef]

40.   Rowan, D.J.; Logunova, V.; Oshima, K. Measured residual tumor cellularity correlates with survival in neoadjuvant treated pancreatic ductal adenocarcinomas. *Ann. Diagn. Pathol.* **2019**, *38*, 93–98. [CrossRef]

41.   Matsuda, Y.; Ohkubo, S.; Nakano-Narusawa, Y.; Fukumura, Y.; Hirabayashi, K.; Yamaguchi, H.; Sahara, Y.; Kawanishi, A.; Takahashi, S.; Arai, T.; et al. Objective assessment of tumor regression in post-neoadjuvant therapy resections for pancreatic ductal adenocarcinoma: Comparison of multiple tumor regression grading systems. *Sci. Rep.* **2020**, *10*, 18278. [CrossRef]

42.   Ren, P.; Xiao, Y.; Chang, X.; Huang, P.-Y.; Li, Z.; Chen, X.; Wang, X. A survey of deep active learning. *arXiv* **2020**, arXiv:2009.00236.

43.   Turkki, R.; Linder, N.; Kovanen, P.; Pellinen, T.; Lundin, J. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *J. Pathol. Inform.* **2016**, *7*, 38. [CrossRef]

44.   Aung, K.L.; Fischer, S.E.; Denroche, R.E.; Jang, G.-H.; Dodd, A.; Creighton, S.; Southwood, B.; Liang, S.-B.; Chadwick, D.; Zhang, A.; et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. *Clin. Cancer Res.* **2018**, *24*, 1344–1354. [CrossRef] [PubMed]

45.   Echle, A.; Rindtorff, N.T.; Brinker, T.J.; Luedde, T.; Pearson, A.T.; Kather, J.N. Deep learning in cancer pathology: A new generation of clinical biomarkers. *Br. J. Cancer* **2021**, *124*, 686–696. [CrossRef] [PubMed]

46.   Calderaro, J.; Kather, J.N. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. *Gut* **2021**, *70*, 1183–1193. [CrossRef] [PubMed]