

The International Radiomics Platform – An Initiative of the German and Austrian Radiological Societies

First Application Examples

Die Internationale Radiomics-Plattform – eine Initiative der Deutschen und Österreichischen Röntgengesellschaften

Erste Anwendungsbeispiele

Authors

Daniel Overhoff¹, Peter Kohlmann², Alex Frydrychowicz³, Sergios Gatidis⁴, Christian Loewe⁵, Jan Moltz², Jan-Martin Kuhnigk², Matthias Gutberlet⁶, H. Winter⁷, Martin Völker⁸, Horst Hahn², Stefan O. Schoenberg¹

Vorstandskommission Radiomics und Big data:

U. Attenberger, F. Bamberg, B. Baeßler, J. Barkhausen, M. Forsting, J. Furtner-Srajer, S. Gatidis, M. Gutberlet, H. K. Hahn, C. Herold, A. Kaindl, J. Kather, R. Kikinis, P. Kohlmann, B. J. Krause, G. Krombach, M. E. Ladd, G. Langs, S. Lohwasser, J. Lotz, D. Pinto dos Santos, H. Prosch, S. O. Schoenberg, M. Völker, H. Winther, M. Zimmermann

Vorstand der Deutschen Röntgengesellschaft:

G. Antoch, F. Anton, J. Barkhausen, A. Dörfler, F. Körber, G. Layer, S. Neumann, S. O. Schoenberg, J. Weßling, M. Wucherer

Präsidium der Österreichischen Röntgengesellschaft:

R. Forstner, E. R. Gizewski, K. A. Hausegger, G. Heinz, C. Herold, C. Loewe, H. Prosch, E. Sorantin, K. Wicke

Affiliations

- 1 Department of Radiology and Nuclear Medicine, University Medical Center Mannheim, Medical Faculty Mannheim, Heidelberg University, Germany
- 2 Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany
- 3 Department of Radiology and Nuclear Medicine, University Hospital Schleswig-Holstein, Campus Lübeck, Germany
- 4 Department of Diagnostic and Interventional Radiology, University-Hospital Tübingen, Germany
- 5 Department of Radiology, Medical University of Vienna, Austria
- 6 Department of Diagnostic and Interventional Radiology, Leipzig Heart Centre University Hospital, Leipzig, Germany
- 7 Department of Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany
- 8 German Roentgen Society „Deutsche Röntgengesellschaft“, Berlin, Germany

Key words

artificial intelligence, myocarditis, magnetic resonance imaging, neural networks

received 02.06.2020

accepted 17.07.2020

published online 2020

Bibliography

Fortschr Röntgenstr

DOI 10.1055/a-1244-2775

ISSN 1438-9029

© 2020. Thieme. All rights reserved.

Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence

Dr. Daniel Overhoff

University Medical Center Mannheim, Department of Radiology and Nuclear Medicine, Theodor-Kutzer-Ufer 1–3, 68167 Mannheim, Germany

Tel.: ++49/6 21/3 83 20 67

daniel.overhoff@umm.de

ABSTRACT

Purpose The DRG-ÖRG IRP (Deutsche Röntgengesellschaft-Österreichische Röntgengesellschaft international radiomics platform) represents a web-/cloud-based radiomics platform based on a public-private partnership. It offers the possibility of data sharing, annotation, validation and certification in the field of artificial intelligence, radiomics analysis, and integrated diagnostics. In a first proof-of-concept study, automated myocardial segmentation and automated myocardial late gadolinium enhancement (LGE) detection using radiomic image features will be evaluated for myocarditis data sets.

Materials and Methods The DRG-ÖRG IRP can be used to create quality-assured, structured image data in combination

with clinical data and subsequent integrated data analysis and is characterized by the following performance criteria: Possibility of using multicentric networked data, automatically calculated quality parameters, processing of annotation tasks, contour recognition using conventional and artificial intelligence methods and the possibility of targeted integration of algorithms. In a first study, a neural network pre-trained using cardiac CINE data sets was evaluated for segmentation of PSIR data sets. In a second step, radiomic features were applied for segmental detection of LGE of the same data sets, which were provided multicenter via the IRP.

Results First results show the advantages (data transparency, reliability, broad involvement of all members, continuous evolution as well as validation and certification) of this platform-based approach. In the proof-of-concept study, the neural network demonstrated a Dice coefficient of 0.813 compared to the expert's segmentation of the myocardium. In the segment-based myocardial LGE detection, the AUC was 0.73 and 0.79 after exclusion of segments with uncertain annotation. The evaluation and provision of the data takes place at the IRP, taking into account the FAT (fairness, accountability, transparency) and FAIR (findable, accessible, interoperable, reusable) criteria.

Conclusion It could be shown that the DRG-ÖRP IRP can be used as a crystallization point for the generation of further individual and joint projects. The execution of quantitative analyses with artificial intelligence methods is greatly facilitated by the platform approach of the DRG-ÖRP IRP, since pre-trained neural networks can be integrated and scientific groups can be networked.

In a first proof-of-concept study on automated segmentation of the myocardium and automated myocardial LGE detection, these advantages were successfully applied.

Our study shows that with the DRG-ÖRP IRP, strategic goals can be implemented in an interdisciplinary way, that concrete proof-of-concept examples can be demonstrated, and that a large number of individual and joint projects can be realized in a participatory way involving all groups.

Key Points:

- The DRG-ÖRG IRP is a web/cloud-based radiomics platform based on a public-private partnership.
- The DRG-ÖRG IRP can be used for the creation of quality-assured, structured image data in combination with clinical data and subsequent integrated data analysis.
- First results show the applicability of left ventricular myocardial segmentation using a neural network and segment-based LGE detection using radiomic image features.
- The DRG-ÖRG IRP offers the possibility of integrating pre-trained neural networks and networking of scientific groups.

Citation Format

- Overhoff D, Kohlmann P, Frydrychowicz A et al. The International Radiomics Platform – An Initiative of the German and Austrian Radiological Societies. *Fortschr Röntgenstr* 2020; DOI 10.1055/a-1244-2775

ZUSAMMENFASSUNG

Ziel Die DRG-ÖRG-IRP (Deutsche Röntgengesellschaft-Österreichische Röntgengesellschaft Internationale Radiomics-Plattform) stellt eine web-/cloudbasierte Radiomics-Plattform auf Grundlage einer öffentlich-privaten Partnerschaft dar. Sie bietet die Möglichkeit der gemeinsamen Nutzung von Daten, Annotation, Validierung und Zertifizierung auf dem Gebiet der künstlichen Intelligenz, Radiomics-Analyse und integrierter Diagnostik. In einer ersten Proof-of-Concept-Studie soll die automatisierte Myokardsegmentation sowie die automatisierte myokardiale Late-Gadolinium-Enhancement (LGE)-Detektion mittels radiomischer Bildmerkmale für Myokarditis-Datensätze evaluiert werden.

Material und Methoden Die DRG-ÖRP-IRP kann zur Erstellung qualitätsgesicherter, strukturierter Bilddaten in Kombination mit klinischen Daten und anschließender integrierter Datenanalyse genutzt werden und zeichnet sich durch die folgenden Leistungskriterien aus: Nutzungsmöglichkeit multizentrischer vernetzter Daten, automatisiert berechnete Qualitätsparameter, Bearbeitung von Annotationsaufgaben, Konturerkennung mittels herkömmlicher Verfahren sowie Verfahren der künstlichen Intelligenz und der Möglichkeit einer gezielten Einbindung von Algorithmen.

In einer ersten Studie wurde ein anhand kardialer CINE-Datensätze vortrainiertes neuronales Netz zur Segmentierung von PSIR-Datensätzen evaluiert. In einem zweiten Schritt wurden radiomische Bildmerkmale zur segmentalen Detektion von LGE der gleichen Datensätze, welche multizentrisch über die IRP zu Verfügung gestellt wurden, angewendet.

Ergebnisse Erste Ergebnisse zeigen die Vorteile (Datentransparenz, Zuverlässigkeit, breite Einbindung aller Mitglieder, kontinuierliche Evolution sowie Validierung und Zertifizierung) dieses plattformbasierten Ansatzes auf. In der Proof-of-Concept-Studie konnte im Vergleich zur Segmentierung des Myokards durch den Experten das neuronale Netzwerk einen Dice-Koeffizienten von 0,813 nachweisen. Im Rahmen der segmentbasierten myokardialen LGE-Detektion ergab sich ein AUC von 0,73 bzw. ein AUC von 0,79 nach Ausschluss von Segmenten mit unsicherer Annotation.

Die Auswertung und Bereitstellung der Daten findet auf der IRP unter Berücksichtigung der FAT-Kriterien (Fairness, Accountability, Transparency) sowie FAIR-Kriterien (findable, accessible, interoperable, reusable) statt.

Schlussfolgerung Es konnte gezeigt werden, dass die DRG-ÖRP-IRP als Kristallisationspunkt für die Generierung weiterer Einzel- und Verbundprojekte genutzt werden kann. Die Durchführung von quantitativen Analysen mit Verfahren der künstlichen Intelligenz wird hierbei durch den Plattformansatz der DRG-ÖRP-IRP erheblich erleichtert, da vorab trainierte neuronale Netze integriert und wissenschaftliche Gruppen vernetzt werden können.

In einer ersten Proof-of-Concept-Studie zur automatisierten Segmentierung des Myokards sowie zur automatisierten myokardialen LGE-Detektion konnten diese Vorteile erfolgreich angewendet werden.

Somit zeigt sich, dass sich mittels der DRG-ÖRP-IRP strategische Ziele interdisziplinär umsetzen, konkrete alltagstaugliche Proof-of-Concept-Beispiele aufzeigen sowie möglichst

partizipativ unter Einbindung aller Gruppierungen eine Vielzahl an Einzel- und Verbundprojekten realisieren lassen.

Introduction

A paradigm shift is currently taking place in modern diagnostic imaging. Radiological findings and reports, for example in the case of an abnormality in the lung, previously only described a mass with a suspected diagnosis of lung cancer. Today, machine learning and the extraction of statistical features make it possible to predict mutations and micrometastases [1]. In addition to this sub-level acquired by machine learning and the extraction of statistical features, there is a diagnostic meta-level that allows conclusions regarding treatment response and survival due to interdisciplinary data integration [2, 3]. Oncology is also undergoing a fundamental change in previous diagnostic-therapeutic procedures. In the event of a CT examination showing multiple metastases, a tissue biopsy at a location randomly selected based on accessibility was performed. A treatment decision was then made based on this. Today, it is possible, at least in principle, to determine the probability of the presence of a specific mutation based on expanded radiomics analyses in the entire body. Based on this, metastases identified by radiomics analysis, for example with the highest risk of intratumoral heterogeneity or secondary mutation, can be biopsied in a targeted manner by interventional radiology [4].

The described often automated extraction of clinically relevant qualitative and quantitative biomarkers from medical image data is referred to as radiomics. This term includes various algorithmic procedures, e. g. classic texture analysis methods as well as artificial intelligence and machine learning, a subset of artificial intelligence. The radiomics method is based on the interpretation of medical image data as a data source that far exceeds traditional visual assessment [5].

Reliable radiomics analysis requires reproducible data processing and consistent quality assurance as shown by multiple studies [4]. In addition to the selection of the correct end point, a combination of reliable segmentation, testing of the stability of extracted features and trained models is important. It is also essential to report the algorithms used and the performed internal and external validation including the methodology that was used. The results of scientific studies to date regarding adherence to these radiomics quality criteria are sobering: Only 1 of approx. 20 studies is at least 50% in compliance with the criteria [6]. Thus the additional information acquired from studies published to date must be viewed with caution despite the euphoria regarding algorithmic image analysis. This criticism is supported by current studies on the robustness of radiomics methods, particularly texture analysis compared to various influencing factors like measurement protocols and reconstruction methods. In addition, the rapid development of methods and knowledge presents a challenge for clinical use [7]. In addition, the clinical routine use of manual or semiautomated segmentation is currently still limited due to the time requirement and interobserver variability.

Deep learning methods of the latest generations show potentially better generalization and adaptivity compared to earlier methods (classic feature extraction in texture analysis) and have the potential to help with the described limitations [8]. The application of deep learning methods requires a comparably large amount of training data, ideally from various locations, and efficient tools for data annotation.

The associated opportunities as well as the significant challenges were the basis for the international radiomics platform project (IRPP) initiated by the German Radiological Society. The IRPP is a cooperative public-private partnership in connection with large university hospitals, Fraunhofer MEVIS, and industry partners in the field of medical and information technology. The goal of this initiative is to generate a not-for-profit, cloud-based analysis platform (International Radiomics Platform, IRP) for the shared use of data, annotation, validation, and certification in the field of artificial intelligence, radiomics analysis, and integrated diagnosis. This is based on a joint consortium agreement analogous to the legal structures for the research campus model of the Federal Ministry of Education and Research [9].

The goal of the study is to describe the structure, features, and possibilities of IRP and to demonstrate the feasibility and initial results based on cardiac MRI application examples.

Methods

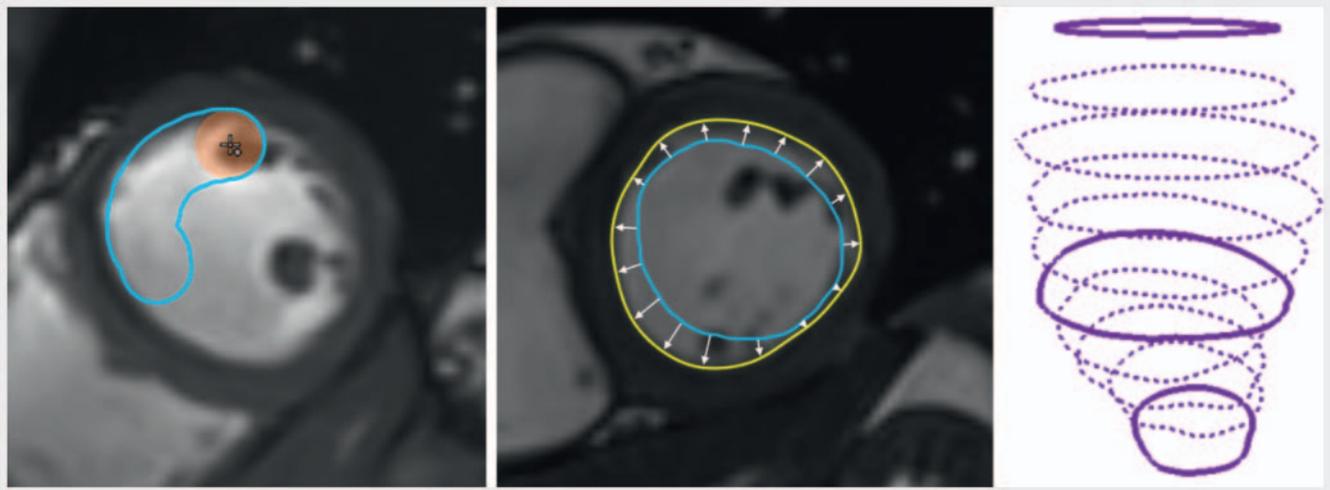
In the currently running DRG-ÖRG IRPP, the web-/cloud-based radiomics platform is used for creating quality-assured and structured image data in combination with clinical data and subsequent integrated data analysis [10]. The retrospective analysis of myocarditis patients with at least two MRI examinations at the time of diagnosis and in the follow-up after six months was used as a pilot project. The maintenance of myocardial function after six months was defined as the end point. At present, six university medicine facilities are participating in the study, all with ethics committee approval, in order to achieve an initial target size of approx. 200 cases.

The IRPP is characterized by the following performance criteria:

1. Provision of a cloud-based platform for the standardized and reliable merging of medical image data and associated information from various locations. At present, web upload of already anonymized image data per drag and drop is available. The IRP will soon be expanded to include integrated anonymization performed by the person doing the uploading. Both de-identification profiles defined in the DICOM standard and configurable positive lists will be supported. The HL7 FHIR standard is to be used for the automated upload of non-image data where possible.



► **Fig. 1** Demonstrator for the visualization and analysis of the quality parameters automatically calculated during data import. The parameters calculated for a selection of image data, such as the signal-to-noise ratio, can be compared with each other in the demonstrator or, if available, with a reference value database.

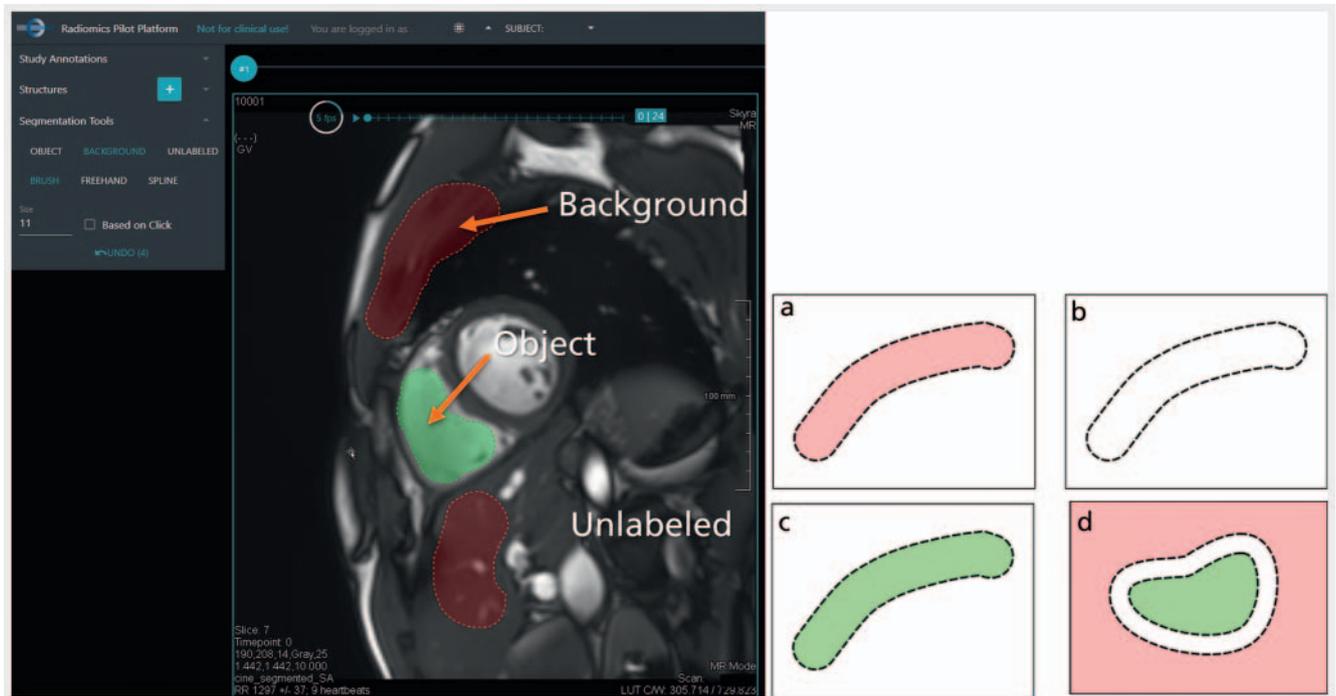


► **Fig. 2** Various tools (e. g. a brush) are available for drawing contours manually (left). The drawn contours can optionally be optimized automatically. A snapping algorithm adapts a drawn contour to high gradients in the environment. The inaccurately drawn blue contour is thereby converted into the yellow contour (middle). If a three-dimensional structure is not drawn on all layers, contours on intermediate layers are supplemented by interpolation (right).

- For the quality analysis of image data, quality parameters regarding signal intensity, homogeneity, and artifacts can be automatically calculated during data import. In a separate web application to be integrated in the IRP, the data sets are compared to one another or to a reference database and are analyzed (► **Fig. 1**). The resulting data can be used to train the algorithm to detect additional artifacts in subsequently uploaded data sets. Moreover, loaded images can be visually assessed by radiology experts (5-star rating in combination with structured input screens).
- Bundling of generic tools for processing a range of annotation tasks. The annotation results are made directly available for radiomics and deep learning (e. g. as training data). The platform is highly configurable and expandable so that it can be

quickly adapted for use in new studies. For efficient manual segmentation of two- and three- dimensional structures, a dedicated toolbox and contouring workflow were implemented. Thus, various tools (Freehand, Spline, Brush) for creating and editing contours and algorithmic support like A) interpolation between contours on layers on which the structure was not manually drawn and B) refining of rough contours using a self-optimizing method (snapping) (► **Fig. 2**) are available. Moreover, innovative interaction concepts were developed to be able to start the training of a neural network with a few high-quality labels using sparse labeling techniques and to iteratively optimize them (► **Fig. 3**).

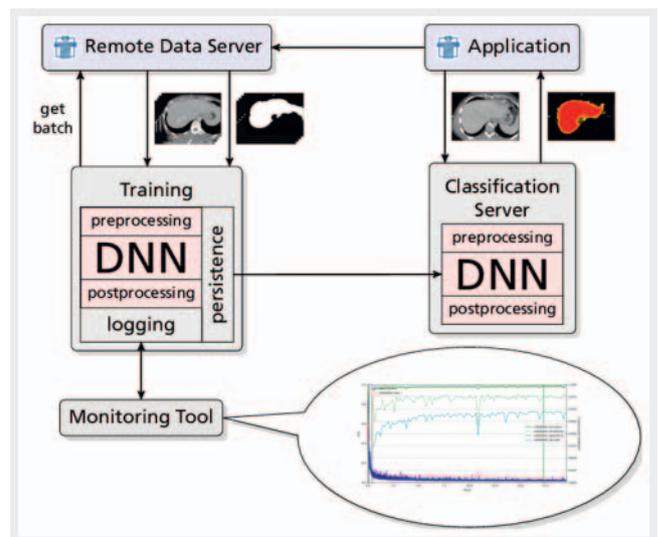
- In addition to the use of conventional contour detection algorithms, special artificial intelligence methods are used to allow



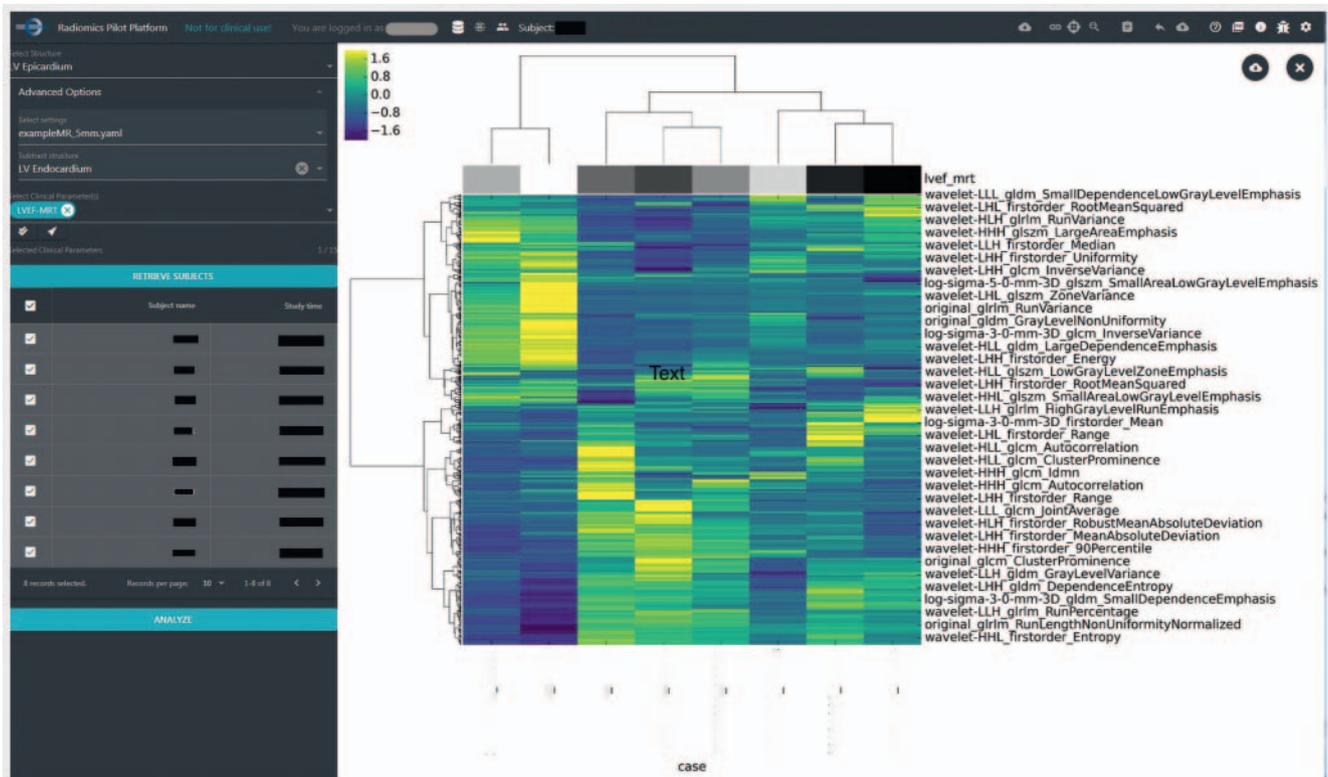
► **Fig. 3** Sparse labeling tool that allows different and incomplete classification of objects in the image: everything within a contour is **a** background, **b** unknown, **c** object, or **d** an uncertainty region. The areas enclosed by the contour are the object, and everything outside the contour is background.

improved segmentation on the basis of already annotated data, e. g. for training neural networks for automated contour analysis of the endocardium and epicardium. For this purpose the radiomics platform is linked to the deep learning framework RedLeaf developed by Fraunhofer MEVIS [11]. The IRP sends newly created or corrected segmentations to a training server for the training of neural networks. A monitoring tool connected to RedLeaf makes it possible to monitor the quality of trained networks. A classification server allows the IRP to retrieve automatic segmentations for structures that already have trained networks (► **Fig. 4**).

5. Targeted inclusion of algorithms is possible, e. g. the use of already trained networks from international algorithm challenges. For this purpose, the IRPP is cooperating with the Grand Challenge Platform [12] which organizes algorithmic comparisons for various medical imaging issues with global participation. The neural networks being used can be further optimized by algorithms already trained with large quantities of data on external servers.
6. Integration of image data and corresponding clinical data as well as clinical end points. Defined end points can be used in combination with image data and clinical data as target variables for modeling statistical models and machine learning methods, e. g. for predicting treatment response.
7. Multidimensional correlations to the corresponding clinical or molecular genetic end points (feature maps) can then be created from the dominant features by machine learning (► **Fig. 5**).



► **Fig. 4** Connection of an application (e. g. IRP) to a training server including a monitoring tool and to a classification server. The training server receives original image data together with corresponding segmentation masks, which were created in the IRP. After receiving these data, training can be started or an already trained network (DNN) can be trained interactively. The monitoring tool connected to the training server monitors the classification accuracy of the trained network in relation to validation data sets. The classification server allows the creation of segmentations in real time from the IRP based on the trained networks.



► **Fig. 5** IRP radiomics analysis: For a selected structure (myocardium: volume within the epicardium minus volume within the endocardium) and selected cases, radiomic features are calculated. A heat map displays these features and, if necessary, further clinical parameters or parameters calculated in IPR (here exemplary *lvef_mrt*: left ventricular ejection fraction) after clustering. Currently, these results can be downloaded for further analysis.

These performance criteria were used in initial studies on automatic segmentation of late gadolinium enhancement (LGE) sequences and in a second step for detecting LGE via radiomic image features.

Current projects

Two multicenter analyses are currently in progress with the support of the German Radiological Society:

1. One MRI study regarding the prediction of cardiac function results (► **Fig. 6a, b**). Data carefully curated by experts are used as the basis for targeted radiomics analyses and the (further) development of deeper neural networks. Multicenter data are evaluated in consensus in this study.
2. One MRI study regarding the evaluation of normal and pathological changes in the wrist at 7 T compared to 3 T (► **Fig. 7**). The main focus of this study is the comparison of radiological images at the two field strengths. In addition to the evaluation of different pathologies, the image quality and presence of image artifacts are assessed visually. The monocentric data are evaluated independently by seven people in this study.

Only an initial analysis of the myocardial MRI study is presented in the following to illustrate the applicability of deep neural networks.

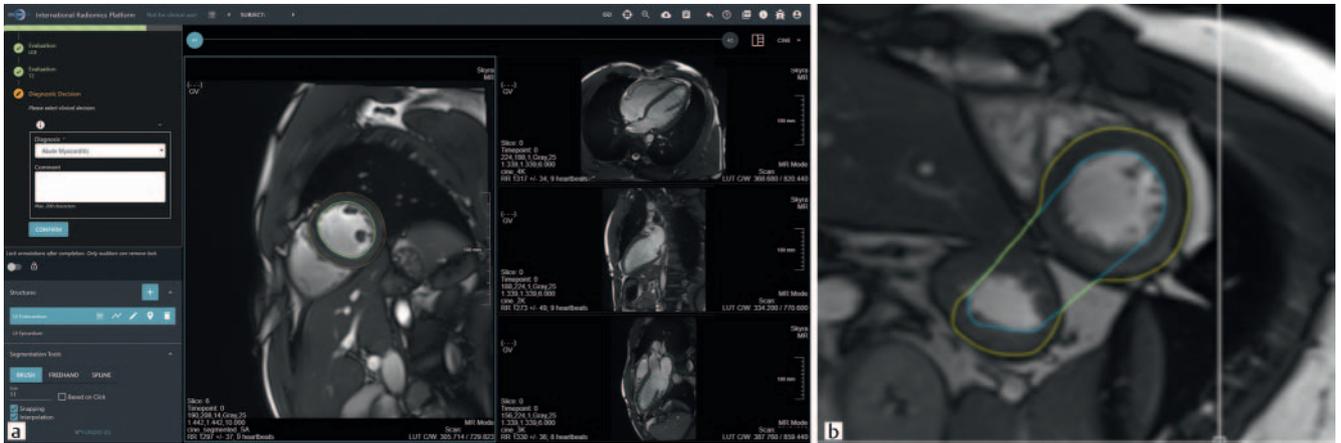
Initial results

47 multicenter cardiac MRI data sets at two time points have been uploaded to the central server for the radiomics analysis. The study was approved by the ethics committees of the participating universities. All 47 patients were included in the initial assessment of the data resulting in a final database of 992 segments to be analyzed (17-segment model of the American Heart Association [AHA]) (► **Fig. 8**).

The LGE images were acquired 10–15 minutes after intravenous administration of gadolinium-containing contrast agent using the “inversion recovery gradient echo” (IR-GRE) pulse sequence. The inversion time (TI) was optimized per patient using a TI scout sequence and was typically between 250 and 300 ms. The phase image of the “phase-sensitive inversion recovery” (PSIR) sequence on the short axis was analyzed for the analysis. The field strength of the MRI scanners was 1.5 and 3 Tesla and the slice thickness of the PSIR sequences was 6–8 mm.

To date, automated segmentation of late enhancement (LGE) data has been established as the first result of the initial application examples. This was performed in two steps: (i) Automated myocardial segmentation via deep learning (2D U-Net architecture [13] with four layers) (► **Fig. 9**) and subsequent (ii) detection of LGE using a Random Forest Classifier.

The LGE (PSIR) data were initially segmented manually by an expert (radiology specialist with more than 5 years of cardiac ima-



▶ **Fig. 6** a IRP configured for the myocarditis study. b CINE-SSFP short-axis view through the left and right ventricle. Example of erroneous segmentation of the initial ML-based algorithm, underlining the need for “supervised learning” even for supposedly simple segmentation tasks. Fluid-filled stomach with thick musculature is misinterpreted by the algorithm as the left ventricular myocardium.

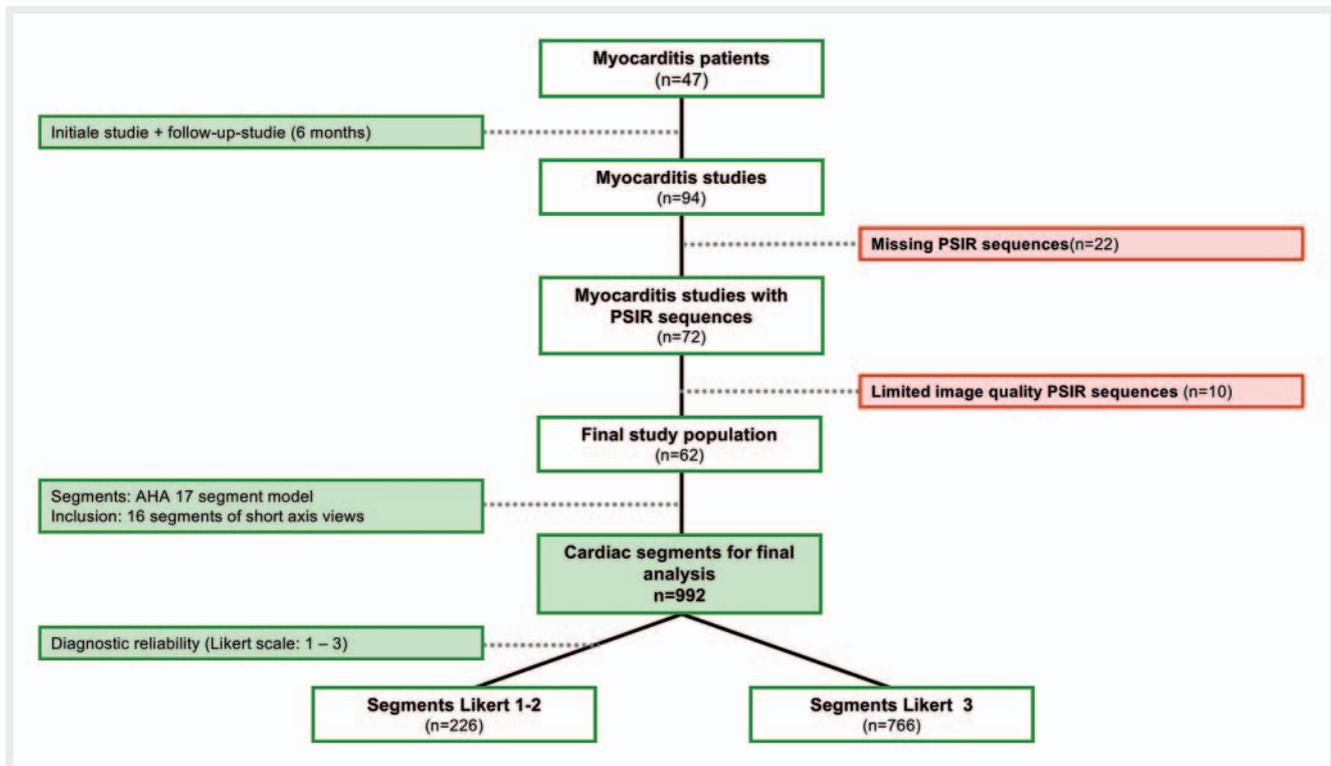


▶ **Fig. 7** IRP configured for the 3T/7T wrist study.

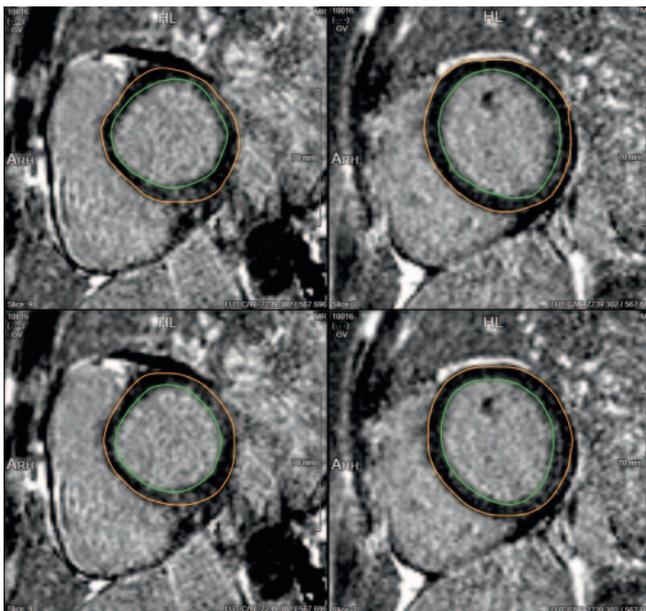
ging experience) to generate training data. Using an existing algorithm for the segmentation of “steady state free precession” (SSFP)-CINE data sets, the neural network was further trained on the basis of the manually segmented LGE data. A neural network that was pre-trained and evaluated with SSFP-CINE data and real time data from 113 patients was used. 80% of 75 data sets from 41 patients (992 frames) were then used to adapt the pre-trained neural network to PSIR data sets. A further 10% of the PSIR data was used for validation and the remaining 10% for testing. This resulted in two classifiers that could be sequentially used for the data. The first classifier identifies an approximate bounding box around the left ventricle and the second classifier segments the myocardium (region between the epicardial and endocardial bor-

der) within this bounding box. The Dice coefficient was determined as a quality criterion for the agreement between the segmentation by the expert and the myocardial segmentation by the neural network. This similarity coefficient indicates both the spatial overlapping and the reproducibility with a Dice coefficient of 1 thus representing complete overlapping or agreement and a value of 0 indicating a lack of agreement [14]. Based on the test data, a Dice coefficient of 0.813 for the myocardium was able to be shown in comparison to the segmentation by the expert. The Dice coefficient for the blood pool (region within the endocardial border) was 0.941.

In the next step the LGE areas were identified by the radiology expert and allocated to the segments of the 17-segment model of



► Fig. 8 Flowchart of the study population with inclusion and exclusion criteria.

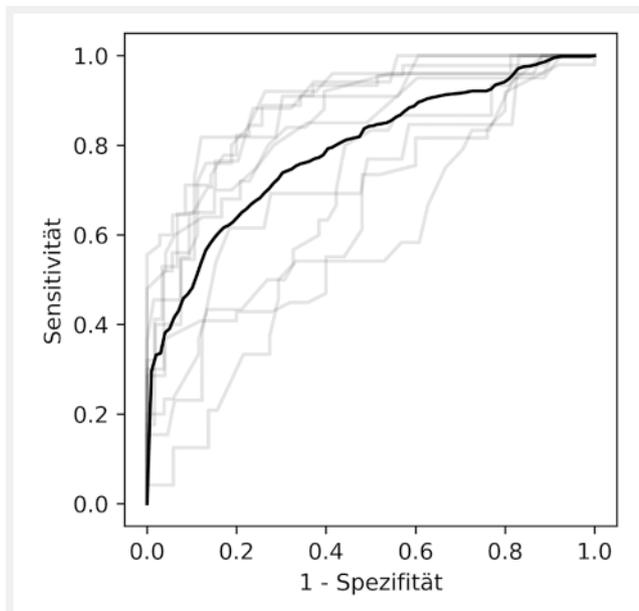


► Fig. 9 Comparison of the segmentation by the expert (upper row) and the DNN (lower row). The data set shown was withheld from the DNN in training. In the left column, a slight pericardial effusion complicates the segmentation.

the AHA classification to be evaluated on the short-axis views and the corresponding segments were classified as LGE positive or negative. The LGE was visually evaluated by the radiology expert

for every segment regarding diagnostic reliability on a three-point Likert scale (1 = low; 2 = average; 3 = high).

Based on this, a study was performed to examine how well late enhancement in individual segments can be detected by radiomic image features. The software library integrated in the IRP (PyRadiomics) [15] was used to calculate a large number of standardized features. Features derived from the intensities and corresponding histograms (features of the first order) and texture features whose calculation included the relationship between multiple voxels (features of a higher order) were included in the analysis. A Random Forest Classifier was used for detecting late enhancement on the basis of these features. The analysis was performed as 10-fold cross-validation with all segments of one patient being assigned to the same group. When using all 992 segments including 408 cases of late enhancement resulted in an average AUC of 0.73. After exclusion of 226 segments with unreliable annotation (points 1 and 2 on a Likert scale of 1 to 3 (► Fig. 8)), the average AUC was 0.79 (ROC curve in ► Fig. 10). At an optimal cut-off value according to the Youden index, this resulted in a sensitivity of 0.62 and a specificity of 0.83. For a sensitivity of 0.8, the specificity would have to be reduced to 0.58. The most important features for this classifier were the average value of the original image and the quantile after application of a Laplacian-of-Gaussian filter to detect edges. Additional experiments showed that prior automated univariate feature selection based on a variance analysis did not provide added value.



► **Fig. 10** ROC curves of the Random Forest Classifier for LGE classification (middle ROC curve bold, ROC curves of the 10-fold cross-validation transparent).

Discussion

The IRP meets the requirements for a radiology society: implementation of strategic goals in a particularly dynamic area in an interdisciplinary manner, concrete proof-of-concept examples and results suitable for the daily routine, and promotion of a number of individual and joint projects in an ideally participatory manner with inclusion of all groups.

The initial use of the IRP for an important clinical question with respect to cardiac imaging shows the significant advantages of this platform-based approach with respect to researching and developing new methods of artificial intelligence:

1. Data transparency: The upload of data, quality assurance, and user access by scientific professional societies ensures high transparency in data analysis using artificial intelligence methods.
2. Reliability: The annotation needed for artificial intelligence methods and thus the “curating” of data can be performed in a multicenter manner by specially selected experts resulting in standardized data sets with the maximum level of quality assurance.
3. Broad inclusion of all members: As a result of the project-based accessibility via the scientific societies, all professional groups with different scientific and clinical interests can define projects and the corresponding end points.
4. Continuous evolution: By including algorithm challenges on the platform, the AI methods can be optimized by the international community in friendly competition.
5. Validation and certification: As a result of the quality-assured control of points 1–4, the scientific societies can validate the algorithms, publish the results on accuracy, reproducibility, and particularly generalizability, and receive certification for clinical use from the corresponding institutions (e. g. TÜV).

This approach also satisfies the FAT criteria (fairness, accountability, and transparency) with respect to ethical aspects when using artificial intelligence for data analysis since data transparency is ensured by the multilateral consortium partnership, the responsibility for the data is ensured by the annotation by appointed experts, and shared availability is ensured by the not-for-profit approach [16]. In particular, the IRPP supports the FAIR criteria by ensuring that the data are consistently findable, accessible, interoperable, and reusable [17]. Particularly regarding interoperability, additional effort, for example in the context of the National Research Data Infrastructure, is needed. A main aspect here is the ability to explain the results in relation to the actual performance and the possible random or systematic deviations resulting from the use of artificial intelligence for specific medical issues. This can be achieved only by precisely defining the basic medical conditions and carefully selecting the resulting end points for testing the neural network.

Cardiac MRI in patients with myocarditis was selected as the first proof-of-concept study. Cardiac MRI is particularly attractive for data analysis using the IRP since it allows the acquisition of standardized, examiner-independent image data and the quantitative calculation of left- and right-ventricular function parameters [18, 19]. This requires segmentation of the endocardium and epicardium in SSFP-CINE sequences. This contouring is time-intensive and examiner-dependent. Semiautomated and automated methods result in significant optimization of the daily workflow as well as further standardization [20–23]. This standardization is an important quality criterion particularly with respect to radiomics analyses [4, 24]. In addition, the clinical picture of myocarditis is particularly suitable for analysis with modern cardiac MRI methods since the quantitative parameters of heart function and inflammatory tissue change in combination with clinical parameters in a standardized, systematic analysis allow a high degree of accuracy with respect to diagnosis and differential diagnosis [25, 26] and statements regarding patient prognosis. In the present study, a neural network was trained and validated with respect to the automated endocardial and epicardial segmentation of PSIR sequences after the application of contrast agent. Manual contouring by a radiology expert served as a reference. The neural network was previously evaluated with a different sequence type, namely SSFP-CINE sequences for cardiac volumetry and cardiac function analysis. The Dice coefficient, a commonly used parameter for comparing the overlapping of segmentations, was calculated to measure agreement.

Previous studies on automated segmentation of the left ventricle usually used CINE sequences. A meta-analysis of earlier “deep learning” neural networks showed an mean Dice coefficient of 0.965 for endocardial left-ventricular contour detection. Isensee et al. achieved the highest Dice coefficient (0.968) among all available studies [27]. The neural network used here yielded similar values. In comparison to threshold-based methods, improved overlapping of the generated segmentation with the “ground truth”, i. e., manual segmentation by an experienced radiologist, was already able to be shown. For example, the values for endocardial segmentation of the left ventricle in SSFP-CINE sequences are 0.88–0.89 [28, 29].

In 2015, Tao et al. were able to achieve a Dice coefficient of 0.81 with an automated segmentation method for contrast-enhanced data sets [30], which is in accordance with value in our study. In general, it must be taken into consideration that cases of myocarditis in which the LGE distribution pattern is typically epicardial or intramyocardial were examined in our study. In comparison to previous studies, this inevitably results in more difficult epicardial contour detection. The lower agreement for the myocardium compared to previous studies is due to the small test population for neural networks. A significant improvement can be expected here particularly because of the international radiomics platform.

In a second approach automated segmental LGE detection (17-segment model of the American Heart Association [AHA]) for PSIR sequences can be analyzed using a Random Forest Classifier. The Random Forest Classifier had an AUC of 0.71 in the ROC curve analysis for segmental LGE detection. Under consideration of the segments that could be evaluated by the expert with high diagnostic reliability using a Likert scale (Likert scale = 3), an improved AUC of 0.77 was seen in the further analysis. Further analyses by the author team are currently targeting the use of the trained neural network and the Random Forest Classifier for the automated diagnosis of myocarditis and differentiation from normal findings and other myocardial pathologies. The systematic annotation of all image data with the corresponding clinical data required for this purpose is currently in the final stage.

A potential advantage of a neural network is that “contamination” by light pixels on the segment border between the endocavitary blood pool and the endocardium as a typical problem in threshold-based methods may be able to be avoided thus possibly resulting in fewer misclassifications of LGE areas. This also seems to be proven by current publications on the topic. Therefore, in a study published in 2019, analysis using deep neural learning has the lowest variance of less than 10% compared to the “ground truth” in Bland-Altman plots, while the variance in the case of threshold-based methods was over 20% [31]. A current publication in Radiology also shows that texture analyses on the basis of already preprocessed data, e. g. with T1 and T2 mapping, allows significantly better classification of patients with myocarditis [32].

To use the IRP as a crystallization point for the generation of further individual and joint projects, routine functionality that makes accumulation, annotation, and evaluation of data as simple as possible and allows continuous and uncomplicated integration of technical innovations must be ensured. The platform approach of the IRP greatly facilitates and accelerates the implementation of quantitative analyses with artificial intelligence methods. There are four main reasons for this:

1. Pre-trained neural networks based on other data sets can be transferred to related issues on the International Radiomics Platform thereby significantly shortening the time needed for further iterative refining to maximum precision (transfer learning).
2. As a result, scientific groups with a similar focus can be networked thereby forming interactive value added chains so that as many scientific questions as possible can be processed while building on each other.

3. (Inter)national networking and the associated increase in data available for the validation of neural networks increases the precision of the networks and allows them to be continuously optimized as required by international professional societies [33]. In addition to the retrospective analysis of already existing data sets, a prospective approach for standardized acquisition and evaluation with respect to the most important parameters to be expected must also be possible.
4. The IRP continues to offer interfaces to other data platforms of large national and international consortia, e. g., the joint imaging platform of the German Cancer Consortium, for targeted radiomics analyses in oncological studies.

To further educate as many young scientists and established radiologists as possible with respect to artificial intelligence methods, a white paper was published by the German Radiological Society based on an interdisciplinary workshop held in 2016 [34]. This requires targeted new alliances between radiologists, computer scientists, mathematicians, and related clinical disciplines [35]. A key component here is the planned junior research academy for training and continuing education in the field of artificial intelligence methods.

CLINICAL RELEVANCE

- The DRG-ÖRG IRP project is a web-/cloud-based radiomics platform based on a public-private partnership.
- The DRG-ÖRG-IRP facilitates, accelerates, and more precisely defines the implementation of quantitative analyses via artificial intelligence.
- Initial study results of the DRG-ÖRG IRP regarding automated myocardial segmentation using a neural network and automated detection of myocardial LGE based on radiomic features show the usability of the platform.
- The DRG-ÖRG IRP allows the transfer of pre-trained neural networks to the platform and networking of scientific groups.

Glossary

Cloud

IT resource for storing or processing data made available remotely in a location-independent manner via Internet or intranet.

Bounding box

Tool for object detection. A square (2D) or a box (3D) is generated around the object to be identified. More precise (automated) segmentation can be performed within the square/box based on this.

Brush

Tool for segmentation/contouring of anatomical structures. Using a graphic circle (brush) with an adjustable diameter, the structure is segmented along the outer brush edge. (► Fig. 2 [left])

Deep learning

Deep learning is a subfield of machine learning using deep neural networks.

Dice

The Dice coefficient (DSC) is a similarity coefficient and indicates both the spatial overlapping and the reproducibility, with a Dice coefficient of 1 representing complete overlapping or agreement and a value of 0 indicating a lack of agreement. The Dice coefficient is calculated as follows: $DSC(A,B) = 2(A \cap B) / (A + B)$ with \cap being the intersection [14].

Feature heat map

Graphic representation of radiomic image features after application of a clustering method. With suitable color coding, the feature heat map provides visual representation of high correlations among radiomic features (and possibly clinical parameters).

Freehand

Tool for segmentation/contouring of anatomical structures. Segmentation is performed via freehand drawing along the structure to be segmented.

Snapping

Correction mode for manually drawn contours. The drawn contour optionally adjusts automatically per algorithm to the significant nearby object borders of the structure to be segmented. (► Fig. 2 [middle])

Cross-validation

Cross-validation refers to a method in which a subset of the data sets is used for training a model and another subset of the data sets is made available for evaluation of the trained network. There are various approaches like K-fold cross-validation or LOOCV (Leave-One-Out Cross-Validation), which are not discussed in greater detail here.

Laplacian-of-Gaussian filter (LoG filter)

This is an image processing filter used for detecting edges.

Sparse labeling technique

This technique can be used to provide local information regarding what is part of the desired structure instead of performing exact contouring of structures. (► Fig. 3)

Random Forest Classifier

This is a machine learning method in which multiple decision trees are generated from subsets of the training data set. The decisions of each individual decision tree are aggregated and the classification with the most votes is defined as the final classification.

PyRadiomics

PyRadiomics is an open-source software package that can be used to extract radiomic features from medical images.

Spline

Tool for segmentation/contouring of anatomical structures. Control points are set along the structure to be segmented. An algorithm interpolates between the control points.

RedLeaf

RedLeaf (Remote Deep Learning Framework). This is a software solution developed by Fraunhofer MEVIS for training, testing, and applying deep neural networks.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] Coroller TP, Grossmann P, Hou Y et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* 2015; 114: 345–350. doi:10.1016/j.radonc.2015.02.015
- [2] Vaidya P, Bera K, Gupta A et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multicohort study for outcome prediction. *The Lancet Digital Health* 2020; 2: e116–e128
- [3] Coroller TP, Agrawal V, Narayan V et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiotherapy and Oncology* 2016; 119: 480–486
- [4] Lambin P, Leijenaar RTH, Deist TM et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017; 14: 749–762. doi:10.1038/nrclinonc.2017.141
- [5] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016; 278: 563–577
- [6] Sanduleanu S, Woodruff HC, de Jong EEC et al. Tracking tumor biology with radiomics: A systematic review utilizing a radiomics quality score. *Radiother Oncol* 2018; 127: 349–360. doi:10.1016/j.radonc.2018.03.033
- [7] Baessler B, Weiss K, Pinto Dos Santos D. Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study. *Invest Radiol* 2019; 54: 221–228. doi:10.1097/rli.0000000000000530
- [8] Hahn HK. Radiomics & Deep Learning: Quo vadis? *Forum* 2020; 35: 117–124. doi:10.1007/s12312-020-00761-8
- [9] Bundesministerium für Bildung und Forschung. Forschungscampus. Im Internet (Stand: 20.05.2020): <https://www.forschungscampus.bmbf.de>
- [10] Klein J, Wenzel M, Romberg D et al. QuantMed: Component-based deep learning platform for translational research. In, *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications: International Society for Optics and Photonics* 2020; 113180U
- [11] MEVIS Fraunhofer. Deep Learning in Medical Imaging. Im Internet (Stand: 20.05.2020): <https://www.mevis.fraunhofer.de/en/solutionpages/deep-learning-in-medical-imaging.html>
- [12] grand-challenge.org. Grand Challenges in Biomedical Image Analysis. Im Internet (Stand: 20.05.2020): <https://grand-challenge.org/>
- [13] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In, *International Conference on Medical image computing and computer-assisted intervention* Springer. 2015: 234–241
- [14] Zou KH, Warfield SK, Bharatha A et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol* 2004; 11: 178–189. doi:10.1016/s1076-6332(03)00671-8
- [15] van Griethuysen JJM, Fedorov A, Parmar C et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017; 77: e104–e107. doi:10.1158/0008-5472.Can-17-0339
- [16] Schönberger D. Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *International Journal of Law and Information Technology* 2019; 27: 171–203

- [17] Wilkinson MD, Dumontier M, Aalbersberg IJ et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*. 2016; 3
- [18] Grothues F, Moon JC, Bellenger NG et al. Interstudy reproducibility of right ventricular volumes, function, and mass with cardiovascular magnetic resonance. *American heart journal* 2004; 147: 218–223
- [19] Maceira A, Prasad S, Khan M et al. Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance* 2006; 8: 417–426
- [20] Avendi M, Kheradvar A, Jafarkhani H. A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical image analysis* 2016; 30: 108–119
- [21] Ngo TA, Lu Z, Carneiro G. Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Medical image analysis* 2017; 35: 159–171
- [22] Codella NC, Weinsaft JW, Cham MD et al. Left ventricle: automated segmentation by using myocardial effusion threshold reduction and intravoxel computation at MR imaging. *Radiology* 2008; 248: 1004–1012
- [23] Mahnken AH, Muhlenbruch G, Koos R et al. Automated vs. manual assessment of left ventricular function in cardiac multidetector row computed tomography: comparison with magnetic resonance imaging. *European radiology* 2006; 16: 1416–1423. doi:10.1007/s00330-006-0226-1
- [24] Zwanenburg A, Vallières M, Abdalah MA et al. The Image Biomarker Standardization Initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; 295: 191145
- [25] Lurz P, Luecke C, Eitel I et al. Comprehensive Cardiac Magnetic Resonance Imaging in Patients With Suspected Myocarditis: The MyoRacer-Trial. *J Am Coll Cardiol* 2016; 67: 1800–1811. doi:10.1016/j.jacc.2016.02.013
- [26] Ferreira VM, Schulz-Menger J, Holmvang G et al. Cardiovascular Magnetic Resonance in Nonischemic Myocardial Inflammation: Expert Recommendations. *J Am Coll Cardiol* 2018; 72: 3158–3176. doi:10.1016/j.jacc.2018.09.072
- [27] Bernard O, Lalande A, Zotti C et al. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? *IEEE Trans Med Imaging* 2018; 37: 2514–2525. doi:10.1109/TMI.2018.2837502
- [28] Huang S, Liu J, Lee LC et al. An image-based comprehensive approach for automatic segmentation of left ventricle from cardiac short axis cine MR images. *J Digit Imaging* 2011; 24: 598–608. doi:10.1007/s10278-010-9315-4
- [29] Liu H, Hu H, Xu X et al. Automatic left ventricle segmentation in cardiac MRI using topological stable-state thresholding and region restricted dynamic programming. *Acad Radiol* 2012; 19: 723–731. doi:10.1016/j.acra.2012.02.011
- [30] Tao Q, Piers SR, Lamb HJ et al. Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. *J Magn Reson Imaging* 2015; 42: 390–399. doi:10.1002/jmri.24804
- [31] Zabihollahy F, White JA, Ukwatta E. Convolutional neural network-based approach for segmentation of left ventricle myocardial scar from 3D late gadolinium enhancement MR images. *Med Phys* 2019; 46: 1740–1751. doi:10.1002/mp.13436
- [32] Baessler B, Luecke C, Lurz J et al. Cardiac MRI and Texture Analysis of Myocardial T1 and T2 Maps in Myocarditis with Acute versus Chronic Symptoms of Heart Failure. *Radiology* 2019; 292: 608–617. doi:10.1148/radiol.2019190101
- [33] Langlotz CP, Allen B, Erickson BJ et al. A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019; 291: 781–791. doi:10.1148/radiol.2019190613
- [34] Deutsche Röntgengesellschaft (DRG) e.V. Radiomics in der Radiologie. Im Internet (Stand: 22.05.2020): <https://www.drg.de/de-DE/3601/radiomics/>
- [35] Recht MP, Dewey M, Dreyer K et al. Integrating artificial intelligence into the clinical practice of radiology: challenges and recommendations. *European radiology* 2020. doi:10.1007/s00330-020-06672-5