



Automated Protocoling for MRI Exams—Challenges and Solutions

Jonas Denck^{1,2,3} · Oliver Haas^{1,2} · Jens Guehring³ · Andreas Maier¹ · Eva Rothgang²

Received: 22 February 2021 / Revised: 13 December 2021 / Accepted: 12 January 2022 / Published online: 30 August 2022
© The Author(s) 2022

Abstract

Automated protocolling for MRI examinations is an amendable target for workflow automation with artificial intelligence. However, there are still challenges to overcome for a successful and robust approach. These challenges are outlined and analyzed in this work. Through a literature review, we analyzed limitations of currently published approaches for automated protocolling. Then, we assessed these limitations quantitatively based on data from a private radiology practice. For this, we assessed the information content provided by the clinical indication by computing the overlap coefficients for the sets of ICD-10-coded admitting diagnoses of different MRI protocols. Additionally, we assessed the heterogeneity of protocol trees from three different MRI scanners based on the overlap coefficient, on MRI protocol and sequence level. Additionally, we applied sequence name standardization to demonstrate its effect on the heterogeneity assessment, i.e., the overlap coefficient, of different protocol trees. The overlap coefficient for the set of ICD-10-coded admitting diagnoses for different protocols ranges from 0.14 to 0.56 for brain/head MRI exams and 0.04 to 0.57 for spine exams. The overlap coefficient across the set of sequences used at two different scanners increases when applying sequence name standardization (from 0.81/0.86 to 0.93). Automated protocolling for MRI examinations has the potential to reduce the workload for radiologists. However, an automated protocolling approach cannot be solely based on admitting diagnosis as it does not provide sufficient information. Moreover, sequence name standardization increases the overlap coefficient across the set of sequences used at different scanners and therefore facilitates transfer learning.

Keywords Machine learning · Magnetic resonance imaging · Automated protocolling · Automated order entry

Introduction

Although the growth of medical imaging utilization has slowed down in comparison with other medical services, radiology services remain crucial for the diagnosis and treatment of many diseases [1, 2]. Claims of overutilization of medical imaging and a high contribution of radiology services to overall healthcare expenditures have led to medical imaging's growth slowdown [3]. However, the contribution of medical imaging to overall healthcare expenditures remains significant.

Consequently, recent discussions have focused on increasing the efficiency and value of medical imaging in general [2] and of specific imaging modalities, such as magnetic resonance imaging (MRI) in particular [4]. Approaches to increase the value of MRI include the development of abbreviated MRI protocols [5, 6], as well as new imaging techniques to accelerate imaging acquisition and increase the diagnostic value. Moreover, artificial intelligence (AI) plays a vital role in developing enhanced imaging techniques and diagnosis support for the radiologist. AI has shown potential to accelerate imaging [7], supports the radiologists during imaging interpretation [8], and is subject to ongoing research to increase efficiency further [9].

Furthermore, the potential of AI applications beyond imaging acquisition and interpretation has already been recognized [10]. It includes, but is not limited to, creating and selecting study protocols, optimizing MR scanner utilization, scheduling, and automated billing [11]. These applications constitute promising approaches to optimize the clinical workflow, reduce the workload for clinical staff and radiologists, and lower costs. It enables radiologists to focus on image interpretation, patient care, and communication.

✉ Jonas Denck
jonas.denck@gmail.com

¹ Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

² Department of Industrial Engineering and Health, Technical University of Applied Sciences Amberg-Weiden, Weiden, Germany

³ Siemens Healthcare GmbH, Erlangen, Germany

However, some of these methods have not reached acceptable performance yet to automate sub-tasks of the clinical workflow fully, which can limit the benefit for the clinical practice.

This is also true for study protocol selection for MRI exams (protocoling). Protocoling describes the process of selecting an adequate imaging protocol under consideration of the ordered procedure, clinical indication, and medical history. A set of pre-defined, institution-specific protocols for each imaging modality is in general available for selection that is listed in the protocol tree of the scanner. Each protocol typically consists of multiple MR pulse sequences, which are parameterized through sequence parameters that determine image contrast, resolution, signal-to-noise ratio, and/or scan time. The parameterized pulse sequences are identified through a user-defined sequence name that describes the resulting contrast and additional information (imaging orientation, pre-/post-contrast, slice thickness, etc.).

Pulse sequences are often used for imaging of multiple body parts and use cases (e.g., turbo/fast spin echo sequences [12]) or can be used almost exclusively for a single body region (e.g., cine sequences used for cardiac MRI [13] or diffusion tensor imaging used for brain MRI [14]). This increases the amount and variability of MR sequences at a scanner and protocoling complexity.

Although protocoling is recognized as an amendable target for workflow automation through the application of AI [15], and efforts are undertaken to automate protocoling (see Section “Literature Review”), published results have several limitations that diminish workflow benefits. These limitations are outlined, analyzed, and quantified in this work.

First, we describe the clinical MRI workflow, including critical workflow steps and potential pitfalls for protocol selection. Then, we review current literature on automated protocoling for MRI exams and perform a data-driven analysis to describe challenges for an AI-based automated protocoling approach for MRI and propose solutions for these challenges.

Our contributions are as follows:

- We provide in-depth analysis of the latest research on automated protocoling for MRI exams and describe common obstacles (information availability, MRI protocol comparability) that restrict its integration into the clinical workflow.
- We analyze the patient’s admitting diagnosis mapped to the assigned MRI protocol and assess the consistency of the mapping.
- We assess the heterogeneity of protocol trees on protocol and sequence level and outline associated problems complicating a robust automated protocoling approach.
- We describe potential solutions to these challenges for successful automated protocoling for MRI exams.

Background

MRI Workflow Description

First, we want to provide a general overview of the clinical workflow for an MRI exam and describe the critical decision steps that impact correct protocoling. The first step for any MRI exam is the referring physician’s procedure order through the transmission of an ordered procedure and the reason for the MRI exam (i.e., clinical indication). The procedure order may require the consultation of a clinical decision support (CDS) system to determine the necessity and appropriateness of the ordering [16].

An insufficient clinical indication or the wrong ordered procedure can be the first source of errors for protocoling. If different MRI systems are available (e.g., 1.5 T and 3 T systems), the schedulers allocate a time slot at the appropriate system for the patient exam. This may be subject to the radiology practice/department’s site conditions as some procedures may only be performed at a specific system, e.g., dedicated systems for a certain body region. The responsible radiologist selects the appropriate protocol depending on the clinical indication, the ordered procedure, prior imaging exams and radiology reports, medical history in the electronic medical record (EMR), and relevant lab results [17]. The technologist performing the exam must select the correct protocol on the MRI host computer as specified by the radiologist. Following exam completion, the radiologist interprets the images and makes a diagnosis. Then, the administrative staff completes documentation and the results are communicated to the referring clinician and the patient.

Protocoling is a time-consuming task and prone to errors [18, 19]. However, it is a crucial workflow step as incorrect protocoling can diminish the clinical value of the subsequent MRI exam. In a workflow observation study, Schemmel et al. measured radiologists’ activities in an academic neuroradiology practice [18]. Within non-image-interpretative tasks (procedures, phone calls, in-room consultation, protocoling, teaching) of radiologists, 17% of their time is spent on protocoling, therefore constituting a significant share of their daily workload (see Fig. 1).

Moreover, Ginat et al. analyzed MRI protocoling errors in the neuroradiology department of a hospital-based academic medical center for 4,244 MRI exams within a 6-month period [19]. In total, 140 protocoling-related issues occurred. They categorized these issues into scheduling errors (1.5 T vs. 3 T), ordering clinician errors (e.g., wrong body region), trainee errors (omitting a crucial sequence), technologist errors (missed pre-processing), and inherent protocol problems (e.g., insufficient field-of-view). 37.1% of the reported errors (scheduling, ordering clinician, trainee errors) are amendable through protocoling automation and workflow improvements.

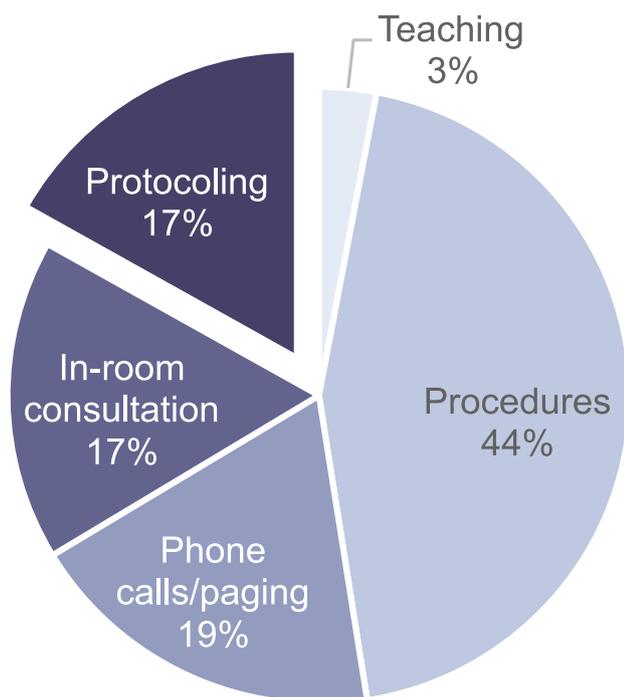


Fig. 1 Timeshare spent by radiologists on different non-image-interpretive tasks [18]

In a former study [20], the impact of missing information within the clinical indication was assessed by evaluating the concordance and completeness of clinical indications in order requisitions compared to provider notes, which offer a more comprehensive picture of the patient's condition. Indications were incomplete in 81% and discordant in 42% within 315 assessed MRI and CT exams. This has a potential impact on correct protocol selection in 8% and on exam interpretation in 43% of the cases.

Protocols form the basis for the patient's diagnosis and further examinations [20]. With an improper protocol, the radiologist might not interpret the MR images accurately, which could lead to patient harm in the worst case. Less severe consequences of incorrect protocols include patient recalls [21] and therefore additional, avoidable costs for the healthcare system. Consequently, every protocoling error must be avoided.

Besides the benefit of avoiding protocoling errors with their associated consequences, automating protocols can result in a significant reduction of the workload for radiologists that can be spent on image interpretation, patient care, and communication. Consequently, automated protocols are anticipated to enhance the clinical MRI workflow considerably.

Literature Review

Different machine learning approaches for automated protocols of MRI exams have been developed recently, with

differences in the underlying machine learning task, input data, and achieved performance, presented in the following (see Table 1).

Brown and Marotta [22] developed three natural language processing (NLP) models that were trained on free-text clinical indications transformed into a term-document matrix, patient and referral demographics to predict the selection of the correct brain MRI protocol (task 1), to evaluate the need for contrast administration (task 2), and to determine priority (task 3) and achieved an accuracy of 82.9%, 83.0%, and 88.2% for tasks 1, 2, and 3.

In another study [23], the authors extended their former study by training a machine learning model to select MRI sequences for the MRI exam based on the input given by patient age and gender, study type, and the study's clinical indication. They used data from a single academic hospital site, where radiologists could choose from 41 different MRI sequences. They reached an accuracy of 95.1% using a gradient boosting machine and state that performing predictions at sequence level allows a more robust clinical application.

IBM Watson was used to predict abdomen/pelvis MRI protocols based on free-text clinical indications [24] and to predict the need for intravenous contrast for musculoskeletal MRI protocols based on free-text clinical indications [25]. They reported an overall accuracy of 83.2% on their test set of 280 cases. However, with respect to a second reader's contrast assignment, who only had access to the clinical indication (without any further information, such as the requested study type), the classifier based on IBM Watson achieved 88.6%. This performance discrepancy shows that the clinical indication alone is sometimes insufficient for determining the need for contrast agent for MRI exams.

A neural network was trained with the word embedding of the clinical indication obtained with Word2vec combined with the requested exam type to predict MRI and CT studies [32, 33]. The network correctly predicted the most likely protocol from over 400 different protocols in 72% and the "top 3" protocol in 90% of the test cases. The error analysis revealed that a large subset of prediction errors was due to data points with minimal clinical, i.e., insufficient information.

The published approaches all use the clinical indication as input as it is the most discriminative feature for protocol selection. In addition to the clinical indication, multiple studies extend the input data by patient demographics, requested procedure, or additional free-text comments. Moreover, different machine learning algorithms have been applied, whereas no specific machine learning model or approach has prevailed and showed superior performance.

Although the presented studies vary in their algorithmic approach, used input data, and achieved performance, they share limitations restricting their use in the daily clinical

Table 1 Publication overview on automated protocolling for MRI exams

Ref	Task	Classification type	Method	Input data	Dataset size (training/ test)	Accuracy ¹
[22]	(1) Protocol selection from 13 MRI brain protocols (2) Contrast administration assignment (3) Priority determination	(1) Multi-class (2) Binary (3) Binary	TDM + random forest	Clinical indication	11,185/2,797	(1) 82.9% (2) 83.0% (3) 88.2%
[23]	Sequence selection from 41 MRI sequences	Multi-label	TDM + gradient boosting machine	Clinical indication, patient age, gender, study type	5,240/2,246	95.1%
[24]	Protocol selection from 14 abdomen/pelvis MRI protocols	Multi-class	IBM Watson	Clinical indication	253/82	93%
[25]	Contrast administration assignment for MSK MRI protocols	Binary	IBM Watson	Clinical indication	1,240/280	83.2%
[26]	Protocol selection from 17 MSK routine and tumor MRI protocols	Multi-class	CNN	Referring department, body region, contrast administration, patient age, gender	5,258/1,018	94.2%
[27]	Protocol selection from 400 CT and MRI protocols	Multi-class	Word2Vec + neural network	Clinical indication Free-text comments Requested procedure	263,700/29,300	72%
[28]	Protocol selection from 108 CT and MRI protocols	Multi-class	Bag of words + neural network	Clinical indication and history	14,519/3,673	84%
[29]	Protocol selection from 88 CT and 293 MRI protocols	Multi-class	TF-IDF + SVM	Requested procedure, department, anatomy, patient history	166,534/41,634 (CT) 392,771/ 98,193 (MRI)	92.2% (CT) 86.9% (MRI)
[30]	Protocol selection from 8 MRI spine and 9 brain protocols	Multi-class	fastText ² + XGBoost ³	Clinical indication	990/151 (spine) 6,067/1,071 (brain)	83.4% (spine) 85.4% (brain)
[31]	Protocol selection from 200/93/48 MRI protocols	Multi-class	One-hot-encoding + neural network	Clinical indication	81,356/34,868	82.8%, 73.8%, 69.3% ⁴

Legend: Method: ^aTDM, term-document matrix; ^bCNN, convolutional neural network; ^cTF-IDF, term frequency–inverse document frequency; ^dSVM, support vector machine. Classification type: ^eBinary: assignment of a binary label (0/1) to a single target variable, e.g., administer contrast medium or not; ^fMulti-class: classification of an instance into one of more than two categories, e.g., select one protocol out of n ($n > 2$); ^gMulti-label: classification of an instance of n different binary labels, e.g., select n MRI sequences for a certain MRI protocol

¹Referring to the top 1 accuracy. Multiple works also evaluated the top 5 accuracy scores

²fastText: library for learning of word embeddings (<https://fasttext.cc/>, accessed 7th November 2021)

³XGBoost: library for gradient boosting (<https://github.com/dmlc/xgboost>, accessed 7th November 2021)

⁴The accuracy scores refer to multiple evaluated protocol subsets

workflow. In the following, we outline and analyze these limitations.

Materials and Methods

In the first part of our work, we assess the information content the clinical indication offers for the protocol selection. Therefore, we analyzed the mapping of ICD-10-coded

admitting diagnosis to MRI protocol and computed the overlap of admitting diagnoses across MRI protocols.

Another challenge for automated protocolling is the heterogeneity of different protocol trees that complicate applying one trained model to a new scanner/protocol tree. We demonstrate that by assessing the protocol and sequence similarity between different protocol trees in the second part of our work.

Little research has been published on the heterogeneity/similarity of MRI protocol trees with respect to the set of

used protocols and sequences [11, 34]. In general, it is difficult to qualitatively compare protocol trees from different scanner types and sites due to the various considerations that must be taken into account for protocol tree configuration. Furthermore, an optimal MRI protocol tree setup can hardly be defined, since external and internal guidelines, types of exams, and application availability (e.g., sequences) vary significantly and change frequently.

However, the similarity of protocol trees can be assessed quantitatively to measure protocol and sequence overlap as a proxy for standardization and harmonization. The degree of similarity of different protocol trees directly impacts whether an automated protocoling approach is transferable, without retraining or fine-tuning, to a new scanner or site.

The methods and used data are described in the following in more detail.

Set Similarity Metrics

For the similarity assessment between the set of admitting diagnoses for different protocols (“[Admitting Diagnosis Protocol Overlap](#)”) and the set of protocols and sequences used across scanners (“[Protocol Tree Assessment](#)”), a suitable similarity metric must be employed. Widely used similarity metrics for two finite sample sets A and B are the Jaccard similarity coefficient (also known as Jaccard index) and the overlap coefficient. The Jaccard index is defined as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

The overlap coefficient (oc) is related to the Jaccard similarity coefficient and defined as follows:

$$oc(A, B) = \frac{|A \cap B|}{\min(|A|, |B|)}$$

However, while both coefficients measure the overlap between two sets, only the overlap coefficient is sensitive to the sizes of the sets. If A is a subset of B (or vice versa), the overlap coefficient is 1.

The overlap coefficient has to be proven to be better suited for small datasets and in particular for different sized sets than the Jaccard index [35].

Admitting Diagnosis Protocol Overlap

We evaluated how consistently the same MRI protocols are assigned to patients with the same clinical indication by computing the overlap coefficient for different protocols with respect to the set of admitting diagnosis ICD-10 codes. This allows us to assess the information share provided by the admitting diagnosis (i.e., clinical indication) for the

protocol decision. We used the admitting diagnosis ICD-10 codes since, unlike the free-text clinical indication, the ICD-10-coded admitting diagnosis constitutes a data basis that enables categorization and comparison of clinical indications. The admitting diagnosis is usually coded during the documentation workflow step and can comprise one or more ICD-10 codes. It corresponds to the referring physician’s free-text clinical indication. Studies evaluating medical coding accuracy show that medical coding data serves as a reliable database for this research [36].

We computed the overlap coefficient with respect to the admitting diagnosis ICD-10 codes for all protocols within a certain body region and the complement overlap coefficient for each protocol. The complement overlap coefficient describes the set of admitting diagnosis ICD-10 codes for a single protocol (denoted by A) with respect to any other protocol in the dataset. More formally, the complement overlap coefficient is denoted by $oc(A, A^c)$, where A^c is the complement of A : $A^c = \{x \in U | x \notin A\}$, with U containing all elements within the (filtered) dataset.

We have compiled a dataset that comprises 2,106 MRI exams across all body regions from three MRI scanners of a single radiology practice (Zwanger-Pesiri Radiology, New York, USA). It contains the ICD-10-coded admitting diagnoses extracted from the institution’s radiology information system and MRI scanner log data [11] that contain information about the MRI exam (performed sequences, acquisition parameters, protocol name). The evaluated data comprise all MRI exams performed with the three scanners within January and June 2019. The data were stripped of any patient-identifiable information.

Protocol Tree Assessment

We also analyzed challenges associated with the heterogeneity of protocols (protocol trees) and assessed the similarity between different protocol trees on protocol and sequence level. Therefore, we computed the overlap coefficient for different protocol trees (for all body regions) and the sequences used within the protocols. However, since the sequence name can be adapted individually, it is not suitable for the objective characterization of a protocol step, i.e., the parameterized MRI sequence. Therefore, different strategies were recently developed to categorize and standardize MRI sequence descriptions, which can be used as a meaningful, standardized feature for machine learning applications [11, 37, 38]. We used a heuristic rule-based approach [11] that generates a standardized sequence name based on the MR acquisition parameters of a sequence (e.g., TR, TE, imaging technique, imaging orientation). We show the effect and benefit of sequence name standardization on assessing sequence heterogeneity across different protocol trees.

Table 2 Data properties of the evaluated MRI exams for the selected body regions

Body region	Spine	Brain/head
Number of exams	727	249
Number of unique ICD-10 codes	127	98
Label cardinality	1.1	1.2
Label density	0.01	0.01
Label diversity (code sets)	177	118

Legend: Label cardinality: average number of admitting diagnosis ICD-10 codes per exam; Label density: cardinality divided by the number of unique codes; Label diversity: number of distinct code sets

We evaluated the protocols from three different MRI scanners of the same radiology practice. Two scanners (scanner 1 and 2) are located at the same site, and one scanner (scanner 3) is located at a different site. All MRI scanners are 3 T MAGNETOM Skyra systems (Siemens Healthcare, Erlangen, Germany) with the same software baseline.

Results

Admitting Diagnosis Protocol Overlap

We analyzed the overlap of admitting diagnosis ICD-10 codes within MRI exams of two selected body regions, brain/head and spine, as MRI is frequently used for various use

cases for these body regions. We only incorporated protocols that were performed more than five times within our dataset.

The average number of admitting diagnosis ICD-10 codes per exam is low for both spine (1.1) and brain/head (1.2), and single code occurrence ranges from 1 to 182 for spine and to 42 for brain/head exams, respectively (Table 2).

Since the frequency of the protocols varies widely within the dataset (ranging from 6 to 341), we computed the overlap coefficient to assess the similarity of admitting diagnosis and protocol mapping.

Figures 2 and 3 visualize the overlap coefficients for the set of admitting diagnosis codes associated with the different brain/head and spine MRI protocols, respectively. As the overlap coefficient is symmetric, i.e., $oc(A, B) = oc(B, A)$, so is the heatmap. Note that A and B comprise here the set of admitting diagnosis code combinations for two different MRI protocols. Thus, an element within these sets can consist of single or multiple ICD-10 codes.

The overlap coefficient with respect to the complement ranges from 0.14 to 0.56 for brain/head MRI protocols and 0.04 to 0.57 for spine protocols. When comparing single protocols, the oc ranges from 0 to 0.56 and 0 to 0.5 for brain/head and spine exams, respectively.

The decision between two protocols can solely be based on the admitting diagnosis only if the overlap coefficient is 0; i.e., no overlap of the associated admitting diagnoses with other protocols exists. Consequently, the lower the overlap coefficient, the better for an automated protocol decision based on the admitting diagnosis.

Fig. 2 Heat map visualization of the overlap coefficient of admitting diagnosis ICD-10 codes between MRI head protocols. The annotated decimal values describe the overlap coefficient for the protocol specified on the y-axis and the protocol specified on the x-axis. The last column denotes the complement overlap coefficient for the protocol specified on the y-axis. Abbreviations:—WITHOUT: no contrast medium is administered. WITH means images were acquired after the administration of contrast medium only. W&WO stands for “with and without” contrast medium. IAC stands for internal auditory canal and TMJ for temporomandibular joint

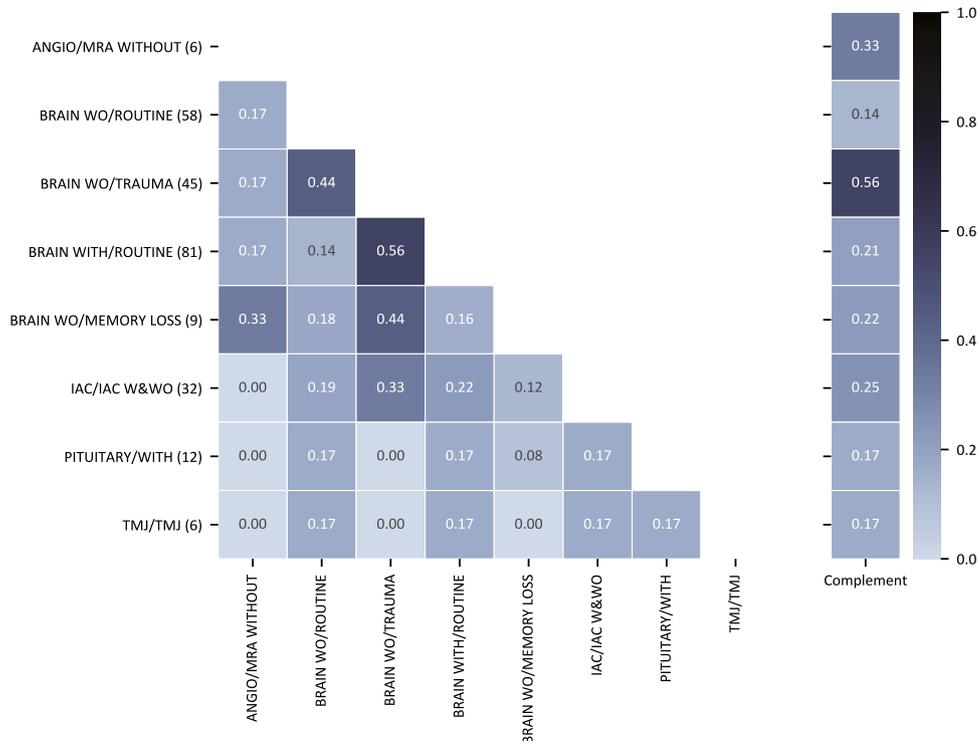
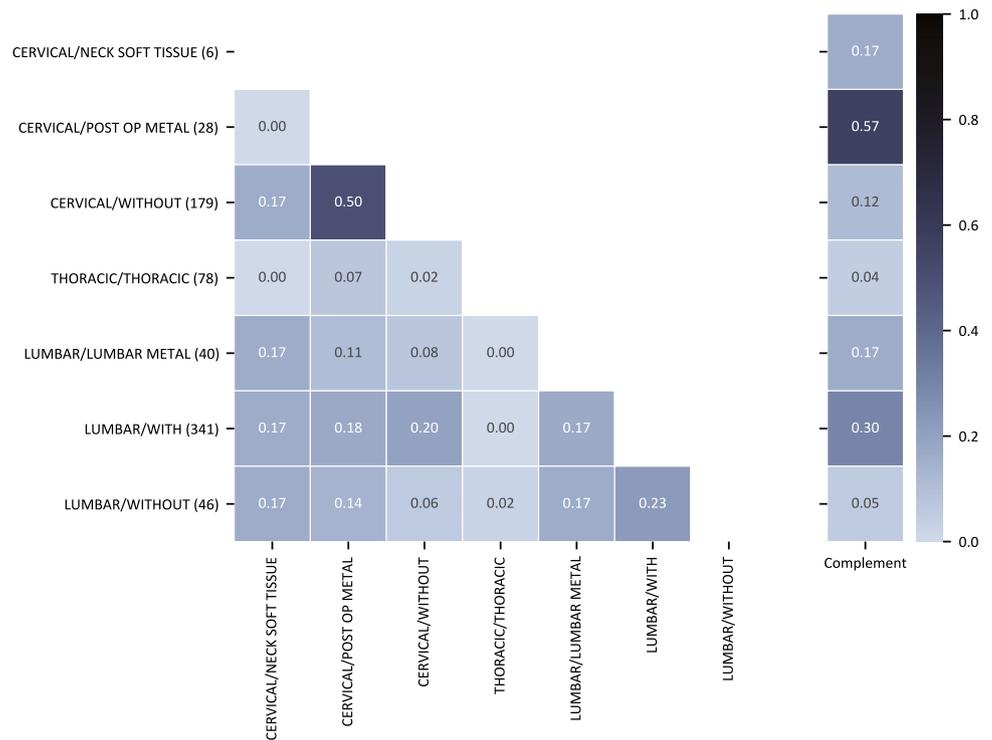


Fig. 3 Heat map visualization of the overlap coefficient of admitting diagnosis ICD-10 codes between MRI spine protocols. The annotated decimal values describe the overlap coefficient for the protocol specified on the y-axis and the protocol specified on the x-axis. The last column denotes the complement overlap coefficient for the protocol specified on the y-axis



Protocol Tree Assessment

We have assessed protocol trees from three scanners of the same model (MAGNETOM Skyra) with the same software baseline, from the same radiology practice, but from two different sites. This allows a valid quantitative comparison of protocol and sequence content and overlap.

On average, 103 protocols are used on each scanner, compiled from 209 different sequence names (across all body regions). MRI sequence categorization and standardization have the advantage that they reduce the variability of sequence descriptions significantly ($P < 0.05$; chi-square test), and sequences become comparable across scanners and sites. When applying sequence name standardization to determine the sequence name solely based on the sequence parameter values, the number of unique sequence names is reduced to 97 on average. In Table 3, the differences in sequence content and protocol naming between the scanners are summarized.

A different set of sequence and protocol names are used on each scanner, with a decreased overlap coefficient for scanners at different sites compared to scanners at the same site. Applying sequence name standardization increases [13] the overlap for the set of sequence names is significantly ($P < 0.05$). Thus, sequence name standardization demonstrates that the difference does not only arise from different sequence content but also from different or inconsistent sequence naming.

The common, i.e., overlapping, set of sequence and protocol names decreases significantly between scanners of the same and different sites. Consequently, an automated

protocolling approach on data from one scanner/site will not be directly applicable to another scanner’s protocol tree, as a different set of sequences requires retraining or fine-tuning.

Discussion

Admitting Diagnosis Protocol Overlap

Through the analysis of the ICD-10-coded admitting diagnosis and the assigned MRI protocol, we showed that MRI protocol selection could not be done reliably on the admitting diagnosis only.

Table 3 Quantitative assessment of the number of different protocols and sequences within the complete protocol trees used at three scanners of the same radiology practice. The overlap coefficient for (standardized) sequence and protocol names is computed with respect to scanner 1

	Scanner 1 (site A)	Scanner 2 (site A)	Scanner 3 (site B)
# Unique sequence names	231	200	196
Unique standardized sequence names	104	98	89
Unique protocol names	107	103	100
<i>oc</i> Sequence names	-	0.86	0.81
Standardized sequence names	-	0.93	0.93
Protocol names	-	0.83	0.79

For example, the *oc* for the spine protocols “CERVICAL/WITHOUT” and “CERVICAL/POST OP METAL” is 0.5; i.e., 50% of the admitting diagnoses associated with “CERVICAL/WITHOUT” were also associated with the “CERVICAL/POST OP METAL” protocol. In this case, crucial information missing is whether metal implants are present in the spine that is not included in the admitting diagnosis. On the other hand, the *oc* for the spine protocols “LUMBAR/WITHOUT” and “LUMBAR/LUMBAR METAL” is significantly lower (0.17). Inconsistencies in protocoling, additional clinical information, or the fact that the protocol name does not reflect the complete application range can lead to these differences.

The presented examples show that multiple protocols for the same admitting diagnosis may be a suitable protocol choice, and more information than only the admitting diagnosis (and consequently also the clinical indication) needs to be assessed to be able to make an informed protocol decision [25]. Selecting the protocol from the clinical indication only could lead to patient harm as contraindications for MRI (e.g., allergies, metal implants, pregnancies) [39] may be overseen, or with less severe consequences, a suboptimal protocol may be selected, and image interpretation may be impeded.

Similar results were obtained for the determination of contrast medium administration [25], as a radiologist assessing the need for contrast medium solely based on the clinical indication achieved only 89.6% accuracy (251/280 cases). Another study identified a significant impact of incomplete exam indications on correct protocoling by analyzing provider notes and clinical indications [20].

However, limitations within our study exist. Although the ICD-10-coded admitting diagnosis does not necessarily reflect the complete clinical indication as information can be lost through coding, it is a useful indicator to assess the information content provided by the clinical indication for the protocol decision. Moreover, it may serve as a systematic approach for protocol configuration by minimizing the overlap with respect to the admitting diagnosis or simply identifying redundant, i.e., overlapping, protocols.

Although our analysis is based on data (2,106 cases) from multiple scanners and sites, it is limited to data from a single radiology practice. Validation of our results with data from additional radiology departments or practices is subject to future work.

Protocol Tree Assessment

In the second part of our work, we computed the overlap coefficient for the set of protocols and sequences used at different MRI scanners. Moreover, we showed that standardized sequence representations lead to a significant reduction

in feature dimensionality across scanners of the same and different sites.

As MRI protocols are usually adapted regularly to reduce protocol creep [34], to be aligned with updated guidelines or after a scanner software update, a comprehensive automated protocoling approach has to cope with protocol changes. However, an automated protocoling approach that has learned to assign the correct protocol name [22, 25, 26] must be retrained whenever new protocols are added or existing protocols are adapted. Therefore, developing a machine learning system to predict required MRI sequences rather than the textual representation of an MRI protocol (i.e., the site-specific protocol name) is crucial to cope with protocol variation and changes and is anticipated to lead to a more robust application. Furthermore, the robustness of an automated protocoling model can be improved with sequence name standardization, which leads to a reduced feature space and makes the approach independent of the user-given sequence name. Despite the protocol heterogeneity of different protocol trees, the aggregation of a large dataset may help to learn different protocoling setups for a certain clinical question and may increase the performance of machine learning models for automated protocoling in future works.

Conclusion

Automated protocoling for MRI exams has the potential to reduce the workload for radiologists and prevent protocoling errors that may diminish the diagnostic value of an MRI exam and lead to re-scans. Unfortunately, recently published approaches have not reached an acceptable performance yet to automate protocol assignment for MRI exams fully and rely only on the clinical indication for the prediction. However, an automated protocoling approach cannot be solely based on the admitting diagnosis as it often does not provide sufficient information for the distinction between two suitable protocols. Moreover, the heterogeneity between protocol trees hinders the deployment of a trained machine learning model to the protocol tree of a new scanner. Sequence name standardization increases the overlap coefficient across the set of sequences used at different scanners and can be crucial to increase the robustness of automated protocoling approaches.

Funding Open Access funding enabled and organized by Projekt DEAL. This work is supported by the Bavarian Academic Forum (BayWISS)—Doctoral Consortium “Health Research,” funded by the Bavarian State Ministry of Science and the Arts.

Availability of Data and Material Private dataset.

Code Availability Custom code.

Declarations

Conflict of Interest Jonas Denck was an employee and Jens Guehring is an employee of Siemens Healthcare GmbH. Andreas Maier has received research grants from Siemens Healthcare GmbH.

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

References

1. D.W. Lee, R. Duszak, D.R. Hughes, Comparative analysis of Medicare spending for medical imaging: sustained dramatic slowdown compared with other services, *Am. J. Roentgenol.* 201 (2013) 1277–1282. <https://doi.org/10.2214/AJR.13.10999>.
2. M.S. Doodoo, R. Duszak, D.R. Hughes, Trends in the utilization of medical imaging from 2003 to 2011: clinical encounters offer a complementary patient-centered focus, *J. Am. Coll. Radiol.* 10 (2013) 507–512. <https://doi.org/10.1016/j.jacr.2013.02.023>.
3. P.-A.T. Duong, B. Bresnahan, D.A. Pastel, G. Sadigh, D. Ballard, J.C. Sullivan, K. Buch, R. Duszak, Value of imaging part I: perspectives for the academic radiologist, *Acad. Radiol.* 23 (2016) 18–22. <https://doi.org/10.1016/j.acra.2015.10.006>.
4. E.J.R. van Beek, C. Kuhl, Y. Anzai, P. Desmond, R.L. Ehman, Q. Gong, G. Gold, V. Gulani, M. Hall-Craggs, T. Leiner, C.C.T. Lim, J.G. Pipe, S. Reeder, C. Reinhold, M. Smits, D.K. Sodickson, C. Tempany, H.A. Vargas, M. Wang, Value of MRI in medicine: more than just another test?, *J. Magn. Reson. Imaging* 49 (2018) e14–e25. <https://doi.org/10.1002/jmri.26211>.
5. C.M. Chhor, C.L. Mercado, Abbreviated MRI protocols: wave of the future for breast cancer screening, *Am. J. Roentgenol.* 208 (2017) 284–289. <https://doi.org/10.2214/AJR.16.17205>.
6. A.S. Chaudhari, M.S. Black, S. Eijgenraam, W. Wirth, S. Maschek, B. Sveinsson, F. Eckstein, E.H.G. Oei, G.E. Gold, B.A. Hargreaves, Five-minute knee MRI for simultaneous morphometry and T2 relaxometry of cartilage and meniscus and for semiquantitative radiological assessment using double-echo in steady-state at 3T, *J. Magn. Reson. Imaging* 47 (2018) 1328–1341. <https://doi.org/10.1002/jmri.25883>.
7. F. Knoll, T. Murrell, A. Sriram, N. Yakubova, J. Zbontar, M. Rabbat, A. Defazio, M.J. Muckley, D.K. Sodickson, C.L. Zitnick, M.P. Recht, Advancing machine learning for MR image reconstruction with an open competition: overview of the 2019 fastMRI challenge, *Magn Reson Med* 84 (2020) 3054–3070. <https://doi.org/10.1002/mrm.28338>.
8. D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, *Annu. Rev. Biomed. Eng.* 19 (2017) 221–248. <https://doi.org/10.1146/annurev-bioeng-071516-044442>.
9. J. Zbontar, F. Knoll, A. Sriram, M.J. Muckley, M. Bruno, A. Defazio, M. Parente, K.J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C.L. Zitnick, M.P. Recht, D.K. Sodickson, Y.W. Lui, fastMRI: an open dataset and benchmarks for accelerated MRI. arXiv:1811.08839 (2018).
10. P. Lakhani, A.B. Prater, R.K. Hutson, K.P. Andriole, K.J. Dreyer, J. Morey, L.M. Prevedello, T.J. Clark, J.R. Geis, J.N. Itri, C.M. Hawkins, Machine learning in radiology: applications beyond image interpretation, *J. Am. Coll. Radiol.* 15 (2018) 350–359. <https://doi.org/10.1016/j.jacr.2017.09.044>.
11. J. Denck, W. Landschütz, K. Nairz, J.T. Heverhagen, A. Maier, E. Rothgang, Automated billing code retrieval from MRI scanner log data, *J. Digit. Imaging.* 32 (2019) 1103–1111. <https://doi.org/10.1007/s10278-019-00241-z>.
12. J. Listerud, S. Einstein, E. Outwater, H.Y. Kressel, First principles of fast spin echo, *Magn. Reson. Q.* 8 (1992) 199–244.
13. N.J. Pelc, R.J. Herfkens, A. Shimakawa, D.R. Enzmann, Phase contrast cine magnetic resonance imaging, *Magn. Reson. Q.* 7 (1991) 229–254.
14. A.L. Alexander, J.E. Lee, M. Lazar, A.S. Field, Diffusion tensor imaging of the brain, *Neurotherapeutics* 4 (2007) 316–329. <https://doi.org/10.1016/j.nurt.2007.05.011>.
15. J. Tudor, C. Klochko, M. Patel, D. Siegal, Order entry protocols are an amenable target for workflow automation, *J. Am. Coll. Radiol.* 15 (2018) 854–858. <https://doi.org/10.1016/j.jacr.2018.02.003>.
16. Appropriate use criteria for advanced diagnostic imaging services: § 414.94, in: Code of Federal Regulations, United States.
17. J. Morey, M., N.M. Haney, W. Kim, Applications of AI beyond image interpretation: Artificial intelligence in medical imaging, in: *Artificial Intelligence in Medical Imaging*, 2019, pp. 129–143.
18. A. Schemmel, M. Lee, T. Hanley, B.D. Pooler, T. Kennedy, A. Field, D. Wiegmann, J.-P.J. Yu, Radiology workflow disruptors: a detailed analysis, *J. Am. Coll. Radiol.* 13 (2016) 1210–1214. <https://doi.org/10.1016/j.jacr.2016.04.009>.
19. D.T. Ginat, P. Uppuluri, G. Christoforidis, G. Katzman, S.-K. Lee, Identification of neuroradiology MRI protocol errors via a quality-driven categorization approach, *J. Am. Coll. Radiol.* 13 (2016) 545–548. <https://doi.org/10.1016/j.jacr.2015.08.027>.
20. R. Lacson, R. Laroya, A. Wang, N. Kapoor, D.I. Glazer, A. Shinagare, I.K. Ip, S. Malhotra, K. Hentel, R. Khorasani, Integrity of clinical information in computerized order requisitions for diagnostic imaging, *J. Am. Med. Inform. Assoc.* 25 (2018) 1651–1656. <https://doi.org/10.1093/jamia/ocy133>.
21. S. Gyftopoulos, D. Kim, E. Aaltonen, L.I. Horwitz, Patient recall imaging in the ambulatory setting, *Am. J. Roentgenol.* 206 (2016) 787–791. <https://doi.org/10.2214/AJR.15.15268>.
22. A.D. Brown, T.R. Marotta, A natural language processing-based model to automate MRI brain protocol selection and prioritization, *Acad. Radiol.* 24 (2017) 160–166. <https://doi.org/10.1016/j.acra.2016.09.013>.
23. A.D. Brown, T.R. Marotta, Using machine learning for sequence-level automated MRI protocol selection in neuroradiology, *J. Am. Med. Inform. Assoc.* 25 (2018) 568–571. <https://doi.org/10.1093/jamia/ocx125>.
24. Jae Ho Sohn, Hari Trivedi, Joseph Mesterhazy, Fouad Al-adel, Thienkhai Vu, Alex Rybkin, Michael Ohliger, Development and validation of machine learning based natural language classifiers to automatically assign MRI abdomen/pelvis protocols from free-text clinical indications, in: *Society of Imaging Informatics in Medicine, Annual Meeting, Pittsburgh, PA, USA, Pittsburgh, PA, USA*, (2017).
25. H. Trivedi, J. Mesterhazy, B. Laguna, T. Vu, J.H. Sohn, Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM Watson's natural language processing algorithm, *J. Digit. Imaging.* 31 (2018) 245–251. <https://doi.org/10.1007/s10278-017-0021-3>.
26. Y.H. Lee, Efficiency improvement in a busy radiology practice: determination of musculoskeletal magnetic resonance imaging protocol using deep-learning convolutional neural networks, *J.*

- Digit. Imaging. 31 (2018) 604–610. <https://doi.org/10.1007/s10278-018-0066-y>.
27. Deep learning for comprehensive automated radiology protocoling, 2018.
 28. A. Kalra, A. Chakraborty, B. Fine, J. Reicher, Machine learning for automation of radiology protocols for quality and efficiency improvement, *J. Am. Coll. Radiol.* 17 (2020) 1149–1158. <https://doi.org/10.1016/j.jacr.2020.03.012>.
 29. P. López-Úbeda, M.C. Díaz-Galiano, T. Martín-Noguerol, A. Luna, L.A. Ureña-López, M.T. Martín-Valdivia, Automatic medical protocol classification using machine learning approaches, *Comput. Methods Programs Biomed.* 200 (2021) 105939. <https://doi.org/10.1016/j.cmpb.2021.105939>.
 30. Y.R. Chillakuru, S. Munjal, B. Laguna, T.L. Chen, G.R. Chaudhari, T. Vu, Y. Seo, J. Narvid, J.H. Sohn, Development and web deployment of an automated neuroradiology MRI protocoling tool with natural language processing, *BMC Med. Inform. Decis. Mak.* 21 (2021) 213. <https://doi.org/10.1186/s12911-021-01574-y>.
 31. A.S. Nencka, M. Sherafati, T. Goebel, P. Tolat, K.M. Koch, Deep-learning based tools for automated protocol definition of advanced diagnostic imaging exams, arXiv preprint arXiv:2106.08963 (2021).
 32. Mikolov Tomas, G.s Corrado, Chen Kai, Dean Jeffrey, Efficient estimation of word representations in vector space, in: International Conference on Learning Representations (ICLR), Scottsdale, Arizona, USA, Scottsdale, Arizona, USA, (2013), pp. 1–12.
 33. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean., Distributed representations of words and phrases and their compositionality, in: *Adv Neural Inf Process Syst*, Lake Tahoe, Nevada, United States, 2013, pp. 3111–3119.
 34. P.B. Sachs, K. Hunt, F. Mansoubi, J. Borgstede, CT and MR protocol standardization across a large health system: providing a consistent radiologist, patient, and referring provider experience, *J. Digit. Imaging.* 30 (2017) 11–16. <https://doi.org/10.1007/s10278-016-9895-8>.
 35. V. Verma, R.K. Aggarwal, A comparative analysis of similarity measures akin to the Jaccard index in collaborative recommendations: empirical and theoretical perspective, *Social Network Analysis and Mining* 10 (2020) 37. <https://doi.org/10.1007/s13278-020-00660-9>.
 36. E.M. Burns, E. Rigby, R. Mamidanna, A. Bottle, P. Aylin, P. Ziprin, O.D. Faiz, Systematic review of discharge coding accuracy, *J. Public Health* 34 (2012) 138–148. <https://doi.org/10.1093/pubmed/fdr054>.
 37. T. Noguchi, D. Higa, T. Asada, Y. Kawata, A. Machitori, Y. Shida, T. Okafuji, K. Yokoyama, F. Uchiyama, T. Tajima, Artificial intelligence using neural network architecture for radiology (AIN-NAR): classification of MR imaging sequences, *Jpn. J. Radiol.* 36 (2018) 691–697. <https://doi.org/10.1007/s11604-018-0779-3>.
 38. R. Pizarro, H.-E. Assemblal, D. de Nigris, C. Elliott, S. Antel, D. Arnold, A. Shmuel, Using deep learning algorithms to automatically identify the brain mri contrast: implications for managing large databases, *Neuroinformatics* 17 (2019) 115–130. <https://doi.org/10.1007/s12021-018-9387-8>.
 39. M. Ghadimi, A. Sapra, Magnetic Resonance Imaging (MRI), Contraindications, in: *StatPearls*, StatPearls Publishing, Treasure Island, FL, USA, (2019).

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