AI-based automated evaluation of image quality and protocol tailoring in patients undergoing MRI for suspected prostate cancer

Kluckert, Jonas ; Hötker, Andreas M ; Da Mutten, Raffaele ; Konukoglu, Ender ; Donati, Olivio F

DOI: https://doi.org/10.1016/j.ejrad.2024.111581

European Journal of Radiology 177 (2024) 111581

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Original Article

AI-based automated evaluation of image quality and protocol tailoring in patients undergoing MRI for suspected prostate cancer

Jonas Kluckert [a,*], Andreas M. Hötker [a], Raffaele Da Mutten [a], Ender Konukoglu [b], Olivio F. Donati [a,c]

[a] Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland
[b] Computer Vision Laboratory, Department of Information Technology and Electrical Engineering, ETH Zurich, Sternwartstrasse 7, 8092 Zurich, Switzerland
[c] Radiology Octorad / Hirslanden, Witellikerstrasse 40, 8032 Zurich, Switzerland

ARTICLE INFO

Keywords:
Multiparametric MRI
Prostate cancer
Artificial Intelligence
Image quality
Personalized medicine

ABSTRACT

Purpose: To develop and validate an artificial intelligence (AI) application in a clinical setting to decide whether dynamic contrast-enhanced (DCE) sequences are necessary in multiparametric prostate MRI.
Methods: This study was approved by the institutional review board and requirement for study-specific informed consent was waived. A mobile app was developed to integrate AI-based image quality analysis into clinical workflow. An expert radiologist provided reference decisions. Diagnostic performance parameters (sensitivity and specificity) were calculated and inter-reader agreement was evaluated.
Results: Fully automated evaluation was possible in 87% of cases, with the application reaching a sensitivity of 80% and a specificity of 100% in selecting patients for multiparametric MRI. In 2% of patients, the application falsely decided on omitting DCE. With a technician reaching a sensitivity of 29% and specificity of 98%, and resident radiologists reaching sensitivity of 29% and specificity of 93%, the use of the application allowed a significant increase in sensitivity.
Conclusion: The presented AI application accurately decides on a patient-specific MRI protocol based on image quality analysis, potentially allowing omission of DCE in the diagnostic workup of patients with suspected prostate cancer. This could streamline workflow and optimize time utilization of healthcare professionals.

1. Introduction

Prostate Cancer (PCa) is the second most diagnosed malignancy in the male population worldwide [1]. Within its diagnostic work-up, magnetic resonance imaging (MRI) has been established as a crucial tool not only for initial diagnosis but also for staging and risk classification [2–7].

A multiparametric MRI protocol, consisting of T2-weighted, diffusion-weighted (DWI) and dynamic contrast enhanced (DCE) sequences [8] is still considered as standard for this examination. More recently, a shortened "biparametric" approach, omitting the DCE sequence, has shown to provide equal diagnostic value in many cases [7,9–11]. Omitting the acquisition of a DCE sequence shortens the examination time and eliminates any risk of contrast agent related side effects [10,12], thereby increasing patient safety, as well as availability and cost-effectiveness of the exam. However, in some cases, the DCE sequence is still valuable to correctly classify a lesion, especially in cases with poor image quality of unenhanced sequences [13].

The challenge lies in identifying which examinations would significantly benefit from an additional contrast-enhanced sequence, as opposed to those for which a shortened protocol would suffice. While having an experienced radiologist evaluate the image quality of each examination in real-time to determine the need for contrast-enhanced sequences would yield optimal protocols for each individual, such an approach would be impractical and inefficient, particularly in large, high-throughput centers, where resource constraints are a significant factor. Assigning the assessment of image quality to radiological technologists (RTs) or less experienced resident radiologists could

Abbreviations: mpMRI, Multiparametric MRI; AI, Artificial Intelligence; PI-RADS, Prostate Imaging Reporting and Data System; DWI, Diffusion-weighted Imaging; DCE, Dynamic contrast-enhanced; CNN, Convolutional Neural Network; API, Application Programming Interface; JWT, Json Web Token; HTTPS, Hypertext Transfer Protocol Secure; CI, Confidence Interval; RT, Radiologic Technologist.

* Corresponding author.
E-mail address: jonas.kluckert@usz.ch (J. Kluckert).

https://doi.org/10.1016/j.ejrad.2024.111581
Received 12 April 2024; Received in revised form 8 June 2024; Accepted 16 June 2024
Available online 17 June 2024

potentially result in reduced accuracy of individualized protocols. In the worst-case scenario, patients might need to return to the radiology department for a second time to acquire additional dynamic contrast-enhanced (DCE) sequences. This would be necessary to complement the acquired unenhanced sequences and obtain a diagnostically satisfactory study.

Previous studies have suggested the use of a convolutional neural network (CNN) to aid in the decision-making process between acquiring a "multiparametric" or a "biparametric" protocol based on T2- and diffusion-weighted images. This approach demonstrated promising results, achieving a sensitivity of 94 % and a specificity of 69 %, leading to a significant reduction in the number of DCE sequences required while maintaining a low patient re-examination rate of only 2 % [14]. However, these findings were obtained in a retrospective study setting after the examinations had already been completed. Our study aimed to validate the benefits of automated decision-making and personalization of prostate MRI in a clinical setting. Specifically, we sought to assess the feasibility and accuracy of automated image analysis and the communication of results to radiology technologists in a prospective manner.

2. Materials and methods

2.1. Workflow

To integrate the CNN into clinical practice and determine the necessity of DCE sequences in real-time, the following workflow (Fig. 1) was established:

A server was set up within the hospital's system to run the application's backend. To communicate the outcomes of automated real-time image analysis to radiologic technologists (RTs), a custom mobile app was developed and provided to RTs and radiologists via mobile devices.

Upon finishing the protocol for obtaining T2-weighted and DWI sequences, the images were transmitted to the server for processing.

After analyzing the image, the results were displayed to the RTs in the mobile app. The results were divided into three categories: "Approved" for images with good quality, "Insufficient" for images with insufficient quality, and "Inconclusive" for cases in which the algorithm could not reach a decision with sufficient certainty. Based on these results, the app recommended the following next steps:

1. Approved: Acquisition is complete and no DCE is needed.
2. Insufficient: Acquisition of a DCE sequence is required.
3. Inconclusive: Consultation with a senior radiologist for a final decision is suggested.

In the first scenario, the RT would merely confirm the acquisition's completion to the mobile app. In the second scenario, the RT would notify the mobile app that a DCE sequence is forthcoming. The case would be marked as finished once this acquisition was complete. In the third scenario, the RT forwarded the case to the supervising radiologist. Subsequently, radiologists would examine cases on their mobile devices that were pending review. The radiologists would enter their decision after examining the images, which would then be sent to the RT's device and processed accordingly.

For this particular implementation, the cutoff values used to classify the CNN's outcome, which represents the likelihood of needing DCE sequences, were chosen as follows: results < 0.3 were labeled as "approved," results > 0.7 were labeled as "insufficient" and required DCE sequences, and results between these cutoffs were labeled as "inconclusive" and required radiologist review. These cutoffs were chosen based on the experience when evaluating the CNN on a validation data set.

2.2. System architecture

The application was organized into a backend and a frontend (Fig. 2.). The backend included a DICOM node, a database (PostgreSQL), analysis scripts, and an API (Application Programming Interface). The API facilitated communication between the mobile client and the backend, adhering to the representational state transfer architectural style, a widely used standard for APIs. The frontend was a mobile client available for iOS and Android. Communication between the frontend

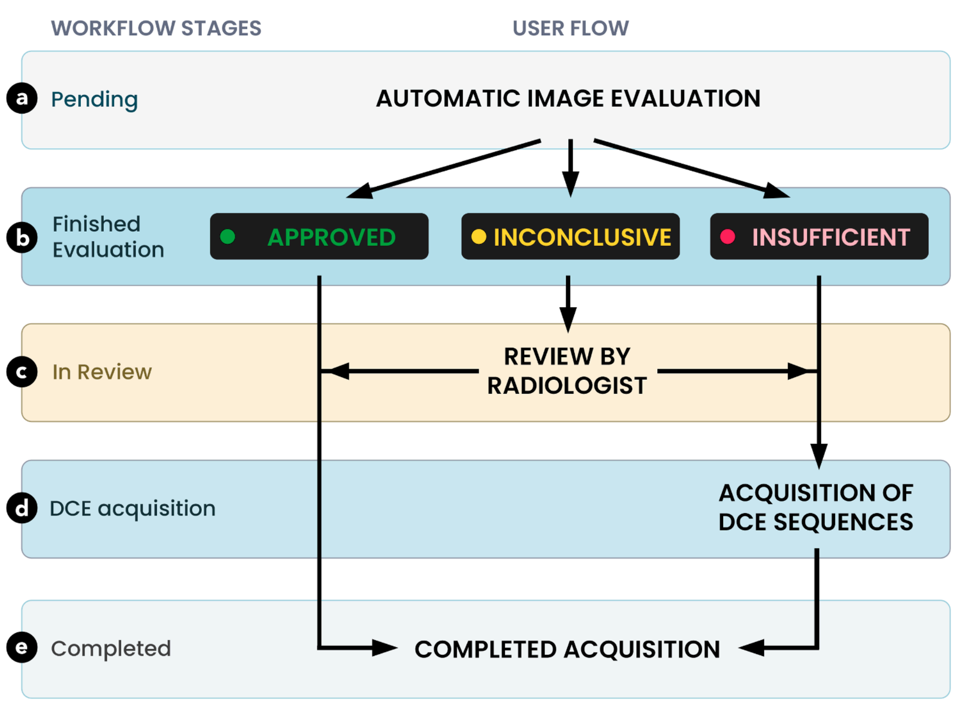

Fig. 1. Workflow: (a) Right after acquisition, the non-contrast sequences are analyzed automatically. (b) The result is classified into one of three labels, approved, inconclusive and insufficient. (c) If inconclusive, the images are sent to an experienced radiologist for review. (d) If insufficient, DCE sequences are acquired. (e) If the non-contrast sequences were approved, no further image acquisition is required and the examination is complete.

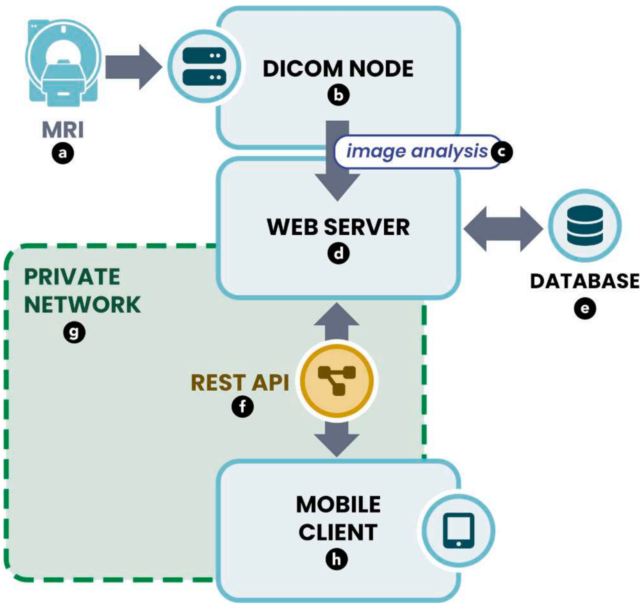

Fig. 2. System architecture: When T2-weighted and DWI sequences are acquired at the MRI (a), they are sent to a DICOM node (b), on which the image analysis algorithm (c) is running. The analysis result is then stored on a web server (d) with a database (e). The mobile client (h), used by the RTs (and radiologists for reviewing images) can poll the analysis results via the REST API (f) of the web server via a secured local network (g).

and backend was secured via the HTTPS protocol within a local network.

Images from the MR scanner were transmitted using the DICOM protocol to the DICOM node of the application. Data transfers were restricted to a controlled network segment. The DICOM node stored and cataloged the received images. Simultaneously, the analysis script regularly checked the DICOM node for incoming files and initiated analysis upon receiving the required images (T2-weighted images in axial, coronal, and sagittal orientations and the DWI sequences).

Image analysis was performed using a validated Convolutional Neural Network (CNN)[14]. The analysis result, including the likelihood of the need for DCE and its interpretation based on predefined cutoff values, was then forwarded to the web server. The server stored the result and metadata about the examination in a database (Table 1). It exposed a REST API to the local network to which the mobile client was connected. Additionally, the server maintained the current workflow stage for each case, ensuring synchronization among all connected clients.

The mobile client was configured for two main user types: RTs and radiologists, by assigning the respective roles. Each role got assigned different permissions to provide appropriate authorization. RTs defaulted to the list of cases that either had a new analysis result or were reviewed by a radiologist. They were allowed to mark cases as completed, as "in DCE" or send them for review by a radiologist. Radiologists in turn were allowed to review cases and mark them as "approved" or "DCE required". All actions were communicated to and

Table 1
Metadata stored during the automatic image evaluation process.

Parameter	*Description*
Acquisition Date Time	Timestamp combining DICOM tags "Study Date" (0008, 0020) and "Study Time" (0008, 0030)
Study Received Date Time	Timestamp of when the T2w- and DWI-sequences have fully been received by the DICOM node.
Result Stored Date Time	Timestamp of the web server after having received the result and having stored it to the database.
Completed Date Time	Timestamp of when the RT marked a case as completed.

confirmed by the server.

2.3. System security

As the application processed sensible patient data, special efforts were dedicated to the system's security and protection of said data. The institution's network security team supported and approved the system design. Industry standard measures were applied, including user authentication and authorization, based on a Json Web Token (JWT) strategy, without the use of third party auth services. JWT is an open standard [15] to securely transmit information, which is commonly used for authentication. Communication was secured by TLS encryption, using the HTTPS protocol. All the user's actions were logged and therefore documented.

2.4. Core libraries

The DICOM node was built using Orthanc [16]. The webserver was a NodeJS application, developed using the Express (v 4.17.1) framework. The script, which was monitoring the DICOM node for any new studies and running the analysis on them, was written in Python (v 3.10.8). The mobile client was developed as a cross-platform app using React Native (v 0.63.4).

2.5. Patients & MRI technique

This study was approved by the institutional review board and the requirement for study-specific informed consent was waived. Between July and December 2022, the system was implemented in our hospital. During this time period, examinations of 53 patients undergoing Prostate MRI were sent to and analyzed by the QRS. One of these cases was excluded, being from a patient who previously underwent prostatectomy. The remaining 52 patients were of a mean age of 64.4 ($\pm$9.7) years.

The MRI examinations were performed on 3 T MRI scanners (Siemens Healthineers, Erlangen, Germany) using MR protocols in compliance with current PI-RADS standards [17].

2.6. Evaluation

To evaluate the QRS in a clinical setting, two additional steps were added to the workflow:

1. The RTs, blinded to the results of the application, gave their own evaluation of image quality.
2. Next, radiology residents blinded to results from application and RTs evaluated image quality.

Both answers were recorded by the app's user interface. Only after entering these two responses, the app would show the algorithm's results. The workflow would then continue as described above.

At our institute, the established procedure was a multiparametric MRI including DCE for every case of suspected Prostate Cancer. During the validation period, all patients were examined according to this standard, regardless of the algorithm's result. However, the RTs and the radiologists used the app as if the system's recommendation on acquiring DCE sequences would have been followed strictly, e.g. marking a case as „completed" after the app recommended not acquiring DCE sequences.

Finally, the performance of the CNN implemented as part of the application, as well as the assessment of RTs and resident radiologists, were evaluated in comparison to a reading of a board-certified radiologist with 14 years of experience in dedicated prostate imaging ('expert radiologist', according to ESUR/ESUI consensus [18]), who was blinded against the examination's reports as well as the decisions of the CNN, RTs and resident radiologists.

2.7. Statistics

The performances of the CNN, the RTs and the residents compared to the reference standard were evaluated by sensitivity and specificity, as well as by assessing the inter-reader agreement of the respective groups with the expert radiologist, using Cohen's Kappa [19]. The latter was interpreted as follows: > 0.75 excellent agreement, 0.59–0.75 good agreement, 0.40–0.58 fair agreement, < 0.40 poor agreement. To assess the impact of the automatic analysis on examination times, the average duration of the software's runtime was logged. All statistics were done using the python programming language (v 3.10.8) with libraries pandas (v 1.5.3), numpy (v 1.23.3) and scikit-learn (v 1.3.0).

3. Results

3.1. Accuracy of the CNN

The software classified 45 of the 52 cases (87 %) as "approved" or "insufficient", leaving 7 (13 %) cases as "inconclusive" and recommended for review. Of 5 examinations that required DCE sequences, 4 cases (80 %) were correctly identified (Fig. 4). No case got a recommendation of DCE sequences that did not need them. 40 of 45 cases (89 %) were correctly classified as sufficient image quality, thus not requiring DCE sequences. 1 (2 %) was falsely marked as sufficient image quality (Fig. 5), thus a return of the patient to acquire the DCE sequences in a second examination would eventually have been necessary. This results in a sensitivity of the algorithm of 0.8 with a specificity of 1.0, and a positive predictive value (PPV) of 1.0 and negative predictive value (NPV) of 0.98. Cohen's Kappa of the software compared to the expert reading for the 45 classified cases was 0.88 (95 % CI 0.76–1.0), demonstrating excellent agreement.

3.2. Comparison with RT and resident's performance

Over all cases, RTs reached a sensitivity of 0.29 and a specificity of 0.98, while resident radiologists showed a sensitivity of 0.29 and a specificity of 0.93. Cohen's Kappa compared to the standard of reference was 0.35 (95 % CI 0.21–0.48) for RTs and 0.25 (95 % CI 0.10–0.39) for residents, both being interpreted as poor agreement. While showing similar performance, Cohen's Kappa between RTs and residents was 0.46, thus showing fair agreement.

3.3. Workflow statistics

The average duration of the automatic evaluation from receiving the images on the DICOM node to having the result available on the API ("Result Stored Date Time" – "Study Received Date Time", see Table 1) was 27 s (SD 1.7 s, maximum of 33 s). This included pre-processing, running the CNN as well as storing the evaluation result to the web server and its database.

4. Discussion

Prostate MRI has become a well-established and often used modality in the diagnostic workup of patients with suspected prostate cancer [2–7]. This raises the need to further optimize the workflow of image acquisition and reporting, while retaining diagnostic image quality. We developed an application to analyze diffusion-weighted (DWI) and T2-weighted images in real-time, informing radiologic technologists of the results while the patient is still in the MRI scanner. This real-time analysis is based on a convolutional neural network (CNN) designed to automatically determine the need for dynamic contrast-enhanced (DCE) sequences or the sufficiency of acquired DWI and T2-weighted images [14]. Our application enables individualized protocol optimization for the current multiparametric MRI (mpMRI) standard, improving the workflow of image acquisition and reporting.

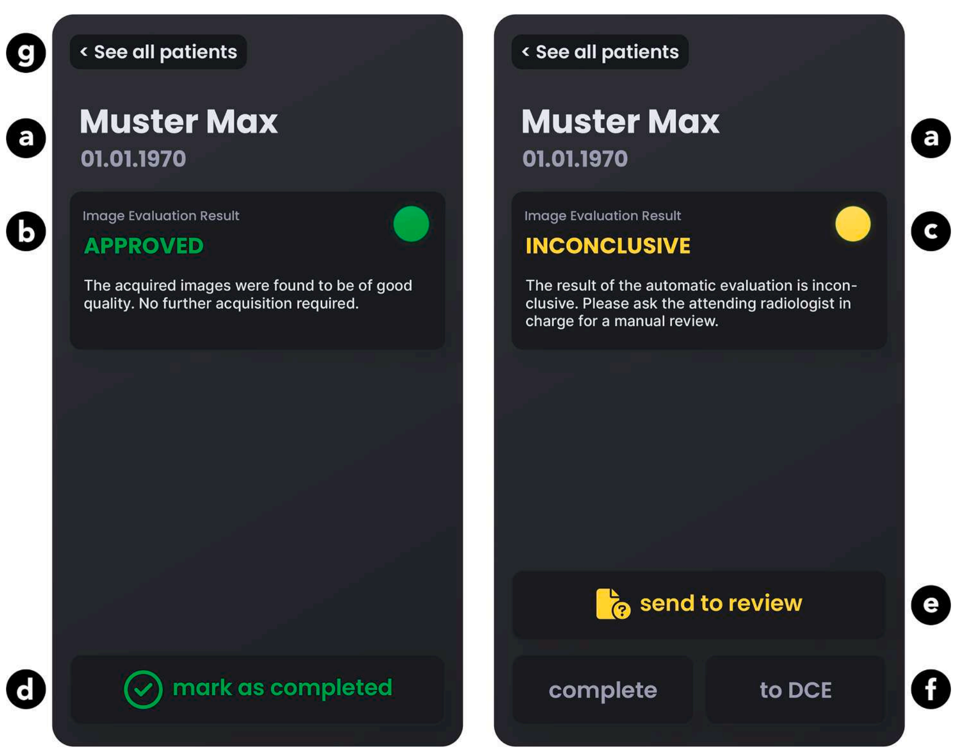

Fig. 3. Mobile App: Two examples of screens displaying detailed information of a single patient to the RTs. (a) Patient's name and birth date for identification. (b) The result of the automated image evaluation, if no DCE sequences are required, and (c) if the result was inconclusive, each with instructions what to do next. (d) Button to mark the current examination as completed, and (e) for sending the case to manual review. (f) gives options to not follow the recommendation and instead complete the examination without DCE sequences or tell the app that DCE sequences will be acquired. (g) Button to return to the list of current patients.

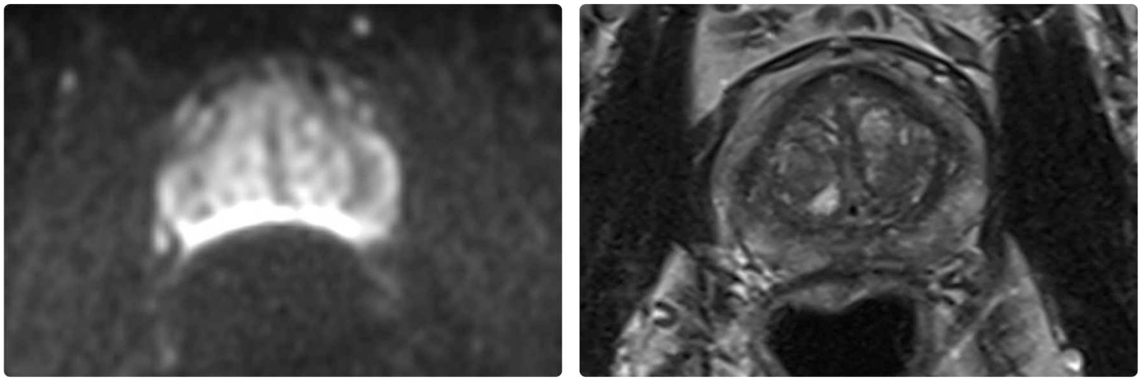

Fig. 4. 57 year-old patient undergoing prostate MRI for suspicion of prostate cancer. Left: diffusion-weighted, b = 1000 s/mm^2. Right: T2-weighted. The AI and expert radiologist agreed on requiring additional DCE sequences for achieving sufficient diagnostic quality.

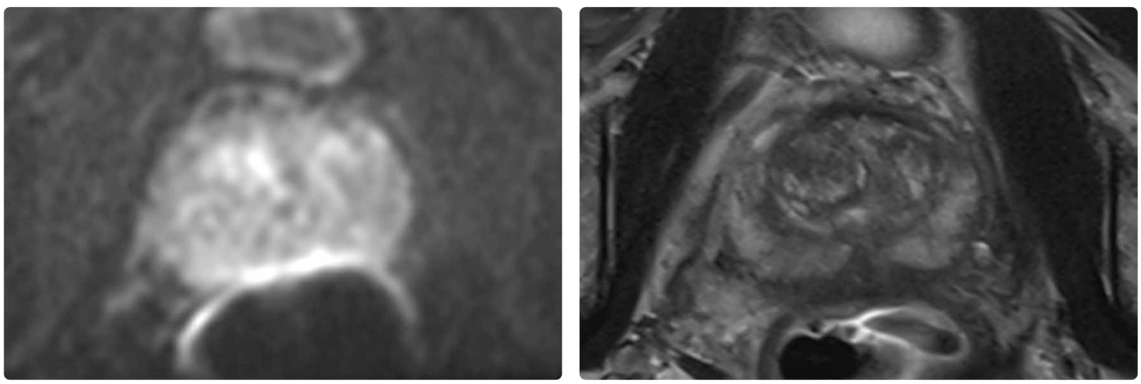

Fig. 5. 74 year-old patient undergoing prostate MRI for suspicion of prostate cancer. Left: diffusion-weighted, b = 1000 s/mm^2. Right: T2-weighted. The AI decided on sufficient image quality. The expert radiologists however deemed DCE sequences necessary, which would have required the patient to return for another examination.

In a clinical setting, we implemented this application during clinically indicated MRI scans of the prostate in patients suspected of having prostate cancer. Although all patients received DCE sequences during this initial trial run, regardless of the application's recommendation, the application showed the potential to significantly reduce (45/52, 87 %) the number of patients who would receive contrast media and additional DCE sequence acquisition. By eliminating the use of intravenous contrast media and acquiring DCE sequences, the patient avoids any risk of adverse effects while benefiting from reduced acquisition time. Moreover, examination costs would be reduced.

An algorithmic threshold was set, prompting machine-automated recommendations in relatively certain scenarios. Consequently, 7 cases (13 %) were classified as "inconclusive" and subsequently forwarded to an expert radiologist for manual evaluation. This process was seamlessly incorporated into the application, permitting real-time display of the review findings to the RTs. This two-tiered approach facilitated excellent diagnostic performance (PPV 1, NPV 0.98), enabling the application to independently manage a substantial portion of cases (45/52, 87 %) without requiring radiologist involvement.

An alternative approach by Rehmann et al. [20] used bi-parametric protocols as a default and recalled patients who required DCE sequences at a rate of 5.7 %. In comparison, the recall rate in our cohort using that system would have been 13.5 % (7/52). If the application's recommendations had been strictly followed, the recall rate would have been reduced to only 1.9 % (1/52).

In comparison to traditional methods, such as manual evaluation by RTs or resident radiologists, the developed application outperformed their recommendations. Notably, the application achieved excellent inter-reader agreement with an expert radiologist (kappa: 0.88), indicating a high level of consistency in image interpretation. In contrast, RTs' and residents' inter-reader agreement with the expert radiologist was relatively poor (kappa: 0.35 and 0.25, respectively). Furthermore, the application demonstrated a significant improvement in sensitivity compared to both technicians and residents. With a sensitivity of 80 % compared to 29 % for both groups, the application increased sensitivity by 51 percentage points. Additionally, by eliminating the need for decision-making in inconclusive cases, the application matches the performance of experts from the original validation of the convolutional neural network (CNN) [14], achieving a Kappa of 0.76. Importantly, the application offers the advantage of reducing human interaction, thus minimizing interruptions to the workflow of radiologists and RTs.

Our CNN-based application was performing well in terms of speed, providing recommendations in approximately half a minute. Nevertheless, at the time of the trial run, the application was still in its early prototype stage, allowing significant room for optimizations. By implementing these enhancements, the duration of the calculations is anticipated to be further reduced. As a result, real-time image analysis for assessing image quality becomes a feasible option in clinical settings without considerably extending examination times.

The application's analysis time cannot be directly compared to the reading time of a human expert radiologist as most time of manual review is taken up by the communication of the RT with the expert radiologist. This time may vary greatly depending on whether the expert radiologist is currently busy with another task. Additionally, interrupting the radiologist during focused work would require them even more time to focus again on the previous task and may increase the risk of errors [21–23].

5. Future directions and unresolved issues

While our applied research showcases the potential of this AI system, several essential considerations and questions remain.

5.1. Trust and confidence

A pivotal aspect to address is the level of trust and confidence that hospital staff have in the AI system's recommendations. Building trust is crucial to ensure the seamless integration of AI-driven decision-making into clinical practice. Further research is necessary to understand the acceptance of such systems, especially in cases where misjudgments occur and how to enhance trust in these applications.

Related measures proposed include:

- Additionally to regulatory approval required for implementing the system beyond research purposes, it would be advisable to carry out standardized testing and validation of the application by an independent party [24].
- Explainable AI: Implement a feature to display areas of interest, e.g. which areas of the images were decisive for the algorithm's result. This would help all stakeholders to understand the AI's decision [25 26].
- Further improvements of the algorithm, especially taking into account that examinations requiring DCE sequences are the minority, but are critical to detect [27]. Also, the decision on which cases should be referred to an expert radiologist for manual review could be made more precisely by including related parameters (confidence in the decision and expert load) into the training of the neural network [28].

5.2. Fully autonomous system

Our current application relies on the availability of an expert radiologist as a fallback option for inconclusive cases. While this setup can reduce the need for personnel resources, a fully autonomous evaluation system would be ideal.

5.3. Scalability and generalizability

The system performed well in the clinical setting, but the trial run was limited to a single institution with MRI scanners from one manufacturer. To facilitate broader adoption, expanding the system to a multi-center setting with machines from different vendors is crucial.

5.4. Limitations of the mobile app

Developing a mobile app was the best approach for this proof-of-concept as it allowed an independent implementation and was easily accessible. Furthermore, it will allow personal push notifications, e.g. to the senior radiologist's device when a review was requested or to the radiologic technologist's device to let them know a patient's case has new information available. Naturally, clients for other platforms could be developed, for example to also display the app's information on the workstation of the MRI machine. An option to store the assessment's result to PACS would be a beneficial addition to the existing application for documentation purposes.

While this "human-in-the-loop" setup is required, one major limitation of the app is that the MRI images cannot be displayed directly in the app, but the senior radiologists must view them on their PACS workstations for review. This will be addressed in future versions to further streamline the review workflow.

6. Conclusion

In conclusion, our research demonstrates the potential of AI in the diagnostic workup of prostate cancer with MRI, specifically in patient tailoring MR protocols by reducing the number of cases requiring DCE sequences when applicable. Our system achieved a performance close to that of an expert radiologist and outperformed residents and RTs. This approach offers a promising way to optimize resource utilization, save time, and further minimize the risk of adverse effects of contrast media.

Ethics

This study was approved by the institutional review board and the requirement for study-specific informed consent was waived.

Funding

None.

Author contributions

All authors contributed to conceptualization and methodology. The software was developed by Jonas Kluckert, Raffaele Da Mutten and Ender Konukoglu. Data collection and curation, as well as analysis were performed by Jonas Kluckert, Andreas Hötker and Olivio Donati. The original draft of the manuscript was written by Jonas Kluckert and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

CRediT authorship contribution statement

Jonas Kluckert: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Andreas M. Hötker:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Raffaele Da Mutten:** Writing – review & editing, Software, Methodology, Conceptualization. **Ender Konukoglu:** Writing – review & editing, Software, Methodology, Conceptualization. **Olivio F. Donati:** .

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Remarks on the development of the mobile app.

The mobile app for iOS (Apple Inc., Cupertino, USA) has been developed with React Native (v 0.63.4), an open source framework based on the React principles. The app consists of three main modules listed below.

- Auth: handles authentication and authorization. When the app is started, users are prompted with a form to enter username and password. Based on the user's role (RT or radiologist) the UI will be altered slightly to show the most relevant information first and to only enable the permitted actions. RT will see the list with the current patients first, radiologists are shown the list of patients pending review. While all users can see all patients of the current day, only radiologists are allowed to review patients.
- Home: displays three lists of patients, sorted by the case's workflow stage. The first list shows patients for which the RT needs to take the next action, e.g. patients who have just received their analysis result. The second list shows patients which require manual review by the radiologist. The third list collects all completed cases of the day. In the home module, the API for analysis results is polled in an interval of 30 seconds.

- Patient Detail: displays the information of one patient including the analysis result. Buttons for user input on how the examination will proceed are available (Fig. 3). This view is accessed through „Home".

The app was developed by a junior radiologist based on the original design and repeated feedback by expert radiologists.

References

[1] M.B. Culp, I. Soerjomataram, J.A. Efstathiou, F. Bray, A. Jemal, Recent global patterns in prostate cancer incidence and mortality rates, Eur. Urol. 77 (2020) 38–52, https://doi.org/10.1016/j.eururo.2019.08.005.

[2] A. O'Shea, M. Harisinghani, PI-RADS: multiparametric MRI in prostate cancer, MAGMA 35 (2022) 523–532, https://doi.org/10.1007/s10334-022-01019-1.

[3] M.C. Fernandes, O. Yildirim, S. Woo, H.A. Vargas, H. Hricak, The role of MRI in prostate cancer: current and future directions, MAGMA 35 (2022) 503–521, https://doi.org/10.1007/s10334-022-01006-6.

[4] V. Kasivisvanathan, A.S. Rannikko, M. Borghi, V. Panebianco, L.A. Mynderse, M. H. Vaarala, et al., MRI-targeted or standard biopsy for prostate-cancer diagnosis, N. Engl. J. Med. 378 (2018) 1767–1777, https://doi.org/10.1056/NEJMoa1801993.

[5] M.C. Velasquez, N.S. Prakash, V. Venkatramani, B. Nahar, S. Punnen, Imaging for the selection and monitoring of men on active surveillance for prostate cancer, Transl Androl Urol 7 (2018) 228–235, https://doi.org/10.21037/tau.2017.08.13.

[6] J.C. Weinreb, J.O. Barentsz, P.L. Choyke, F. Cornud, M.A. Haider, K.J. Macura, et al., PI-RADS prostate imaging - reporting and data system: 2015, version 2, Eur. Urol. 69 (2016) 16–40, https://doi.org/10.1016/j.eururo.2015.08.052.

[7] T. Tamada, A. Kido, A. Yamamoto, M. Takeuchi, Y. Miyaji, T. Moriya, et al., Comparison of biparametric and multiparametric MRI for clinically significant prostate cancer detection with PI-RADS version 2.1, J. Magn. Reson. Imaging 53 (2021) 283–291, https://doi.org/10.1002/jmri.27283.

[8] L. Dickinson, H.U. Ahmed, C. Allen, J.O. Barentsz, B. Carey, J.J. Futterer, et al., Magnetic resonance imaging for the detection, localisation, and characterisation of prostate cancer: recommendations from a European consensus meeting, Eur. Urol. 59 (2011) 477–494, https://doi.org/10.1016/j.eururo.2010.12.009.

[9] B. Wang, J. Gao, Q. Zhang, C. Zhang, G. Liu, W. Wei, et al., Investigating the equivalent performance of biparametric compared to multiparametric MRI in detection of clinically significant prostate cancer, Abdom Radiol (NY) 45 (2020) 547–555, https://doi.org/10.1007/s00261-019-02281-z.

[10] M.H. Choi, C.K. Kim, Y.J. Lee, S.E. Jung, Prebiopsy biparametric MRI for clinically significant prostate cancer detection With PI-RADS Version 2: a multicenter study, AJR Am. J. Roentgenol. 212 (2019) 839–846, https://doi.org/10.2214/AJR.18.20498.

[11] X.-K. Niu, X.-H. Chen, Z.-F. Chen, L. Chen, J. Li, T. Peng, Diagnostic performance of biparametric MRI for detection of prostate cancer: a systematic review and meta-analysis, AJR Am. J. Roentgenol. 211 (2018) 369–378, https://doi.org/10.2214/AJR.17.18946.

[12] M. Scialpi, E. Martorana, P. Scialpi, A. D'Andrea, R. Torre, A. Di Blasi, et al., Round table: arguments in supporting abbreviated or biparametric MRI of the prostata protocol, Abdom Radiol (NY) (2020), https://doi.org/10.1007/s00261-020-02510-w.

[13] J.P. Zawaideh, E. Sala, N. Shaida, B. Koo, A.Y. Warren, L. Carmisciano, et al., Diagnostic accuracy of biparametric versus multiparametric prostate MRI: assessment of contrast benefit in clinical practice, Eur. Radiol. 30 (2020) 4039–4049, https://doi.org/10.1007/s00330-020-06782-0.

[14] A.M. Hötker, R. Da Mutten, A. Tiessen, E. Konukoglu, O.F. Donati, Improving workflow in prostate MRI: AI-based decision-making on biparametric or multiparametric MRI, Insights Imaging. 12 (1) (2021) 112, https://doi.org/10.1186/s13244-021-01058-7.

[15] Jones MB, Bradley J, Sakimura N. JSON Web Token (JWT) 2015. https://datatracker.ietf.org/doc/html/rfc7519 (accessed October 26, 2023).

[16] S. Jodogne, The orthanc ecosystem for medical imaging, J. Digit. Imaging 31 (2018) 341–352, https://doi.org/10.1007/s10278-018-0082-y.

[17] Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. Eur Urol 2019;76:340–51. Doi: 10.1016/j.eururo.2019.02.033.

[18] M. de Rooij, B. Israël, M. Tummers, H.U. Ahmed, T. Barrett, F. Giganti, et al., ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training, Eur. Radiol. 30 (2020) 5404–5416, https://doi.org/10.1007/s00330-020-06929-z.

[19] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174, https://doi.org/10.2307/2529310.

[20] I. Rehman, E. Pang, A.C. Harris, S.D. Chang, Bi-parametric prostate MRI with a recall system for contrast enhanced imaging: Improving accessibility while maintaining quality, Eur. J. Radiol. 169 (2023) 111186, https://doi.org/10.1016/j.ejrad.2023.111186.

[21] S.H. Shah, L.A. Atweh, C.A. Thompson, S. Carzoo, R. Krishnamurthy, N. A. Zumberge, Workflow interruptions and effect on study interpretation efficiency, Curr. Probl. Diagn. Radiol. 51 (2022) 848–851, https://doi.org/10.1067/j.cpradiol.2022.06.003.

[22] N.R. Achangwa, N. Nierobisch, R. Ludovichetti, G. Negrão de Figueiredo, M. Kupka, A. De Vere-Tyndall, et al., Sustainable reduction of phone-call interruptions by 35% in a medical imaging department using an automatic voicemail and custom call redirection system, Curr. Probl. Diagn. Radiol. 53 (2024) 246–251, https://doi.org/10.1067/j.cpradiol.2024.01.004.

[23] B.J. Balint, S.D. Steenburg, H. Lin, C. Shen, J.L. Steele, R.B. Gunderman, Do telephone call interruptions have an impact on radiology resident diagnostic accuracy? Acad. Radiol. 21 (2014) 1623–1628, https://doi.org/10.1016/j.acra.2014.08.001.

[24] J. Ross, S. Hammouche, Y. Chen, A.G. Rockall, Royal college of radiologists AI working group. beyond regulatory compliance: evaluating radiology artificial intelligence applications in deployment, Clin. Radiol. 79 (2024) 338–345, https://doi.org/10.1016/j.crad.2024.01.026.

[25] Brady AP, Allen B, Chong J, Kotter E, Kottler N, Mongan J, et al. Developing, Purchasing, Implementing and Monitoring AI Tools in Radiology: Practical Considerations. A Multi-Society Statement from the ACR, CAR, ESR, RANZCR and RSNA. Radiol Artif Intell 2024;6:e230513. Doi: 10.1148/ryai.230513.

[26] White Paper on Artificial Intelligence: a European approach to excellence and trust - European Commission n.d. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (accessed May 23, 2024).

[27] S. Sangalli, E. Erdil, A. Hötker, O. Donati, E. Konukoglu, Constrained optimization to train neural networks on critical and under-represented classes, Adv. Neural Inf. Proces. Syst. (2021).

[28] S. Sangalli, E. Erdil, E. Konukoglu, Expert load matters: operating networks at high accuracy and low manual effort, Adv. Neural Inf. Proces. Syst. (2023).