



Research article

MRI edema patterns in acute neck infections: A multicenter and multidisciplinary interobserver study

Jari-Pekka Vierula^a, Katri Aro^b, Michaela Bode^c, Eeva Haapio^d, Juhana Hakumäki^e, Tatu Happonen^a, Carita Hollmén^a, Heidi Huhtanen^a, Heikki Irjala^d, Anna Jouppila-Mättö^f, Ville Jussila^a, Kalle Kaapu^g, Jari O. Karhu^a, Petri Koivunen^h, Marcin Kolasa^g, Antti Kovalainen^g, Goran Kurdoⁱ, Eero Kytö^d, Tiina Laurikainen^a, Riikka Lindénⁱ, Ilkka Martikainen^g, Kristofer Nymanⁱ, Timo Paananen^a, Teemu Paavilainen^a, Jarkko Saravo^e, Jutta Tähtinen^g, Ville Tammilehto^a, Jarno Velhonoja^d, Niklas Ånäs^g, Aapo Sirén^a, Jaakko Heikkinen^a, Janne Nurminen^a, Mikko Nyman^a, Kimmo Mattila^a, Jussi Hirvonen^{a,g,*}

^a Department of Radiology, University of Turku and Turku University Hospital, Turku, Finland

^b Department of Otorhinolaryngology-Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

^c Department of Radiology, University of Oulu and Oulu University Hospital, Oulu, Finland

^d Department of Otorhinolaryngology-Head and Neck Surgery, University of Turku and Turku University Hospital, Turku, Finland

^e Department of Radiology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

^f Department of Otorhinolaryngology-Head and Neck Surgery, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

^g Medical Imaging Center, Department of Radiology, Tampere University and Tampere University Hospital, Tampere, Finland

^h Department of Otorhinolaryngology-Head and Neck Surgery, Oulu University Hospital and Medical Research Center, Clinical Medicine, University of Oulu, Oulu, Finland

ⁱ Radiology, HUS Diagnostic Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

ARTICLE INFO

Keywords:

MRI
Neck infections
Edema
Diagnostic accuracy
Interobserver agreement

ABSTRACT

Purpose: To assess the diagnostic accuracy and interobserver agreement of radiologists and surgeons in detecting prognostic MRI edema patterns, retropharyngeal (RPE) and mediastinal (ME) edema, in acute neck infections, supporting their clinical integration into emergency imaging workflows. These patterns are known to predict disease severity, yet their reliable identification across specialties remains untested until now.

Methods: In this multicenter observational study, 28 readers (comprising radiology residents, board-certified radiologists, and head and neck surgeons from five academic hospitals) evaluated axial T2-weighted Dixon MRI images from 60 patients with clinically confirmed acute neck infections. Readers, blinded to clinical data such as patient symptoms or laboratory results, assessed images for RPE and ME presence and rated their confidence. Sensitivity, specificity, accuracy, and interobserver agreement (Fleiss' kappa) were calculated against a consensus standard set by two fellowship-trained neuroradiologists. Intraobserver consistency was tested by repeating 10 randomly selected cases, unbeknownst to readers.

Results: Overall sensitivity, specificity, and accuracy were 0.92, 0.80, and 0.87 for RPE and 0.88, 0.86, and 0.85 for ME, with substantial interobserver agreement ($\kappa = 0.61$ for RPE, 0.52 for ME). Radiologists outperformed surgeons in RPE sensitivity ($p = 0.007$), RPE accuracy ($p = 0.04$), ME sensitivity ($p = 0.01$), and ME accuracy ($p = 0.02$), with intraobserver consistency reaching 90 %. Confidence was high, averaging 4.3 for RPE and 4.1 for ME, and strongly correlated with accuracy ($p < 0.001$), reinforcing diagnostic reliability.

Conclusion: Physicians reliably detect MRI edema patterns across expertise levels, validating their use for risk stratification in acute neck infections and supporting broader adoption in clinical practice where MRI is available.

* Corresponding author at: Department of Radiology, Turku University Hospital and University of Turku, Kiinamyllynkatu 4-8, 20521 Turku Finland.
E-mail address: jussi.hirvonen@utu.fi (J. Hirvonen).

<https://doi.org/10.1016/j.ejrad.2025.112526>

Received 5 August 2025; Received in revised form 29 September 2025; Accepted 16 November 2025

Available online 17 November 2025

0720-048X/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Acute infections of neck soft tissues are a medical emergency marked by a high risk of complications, often requiring prompt diagnosis and treatment [1]. Recent studies have established magnetic resonance imaging (MRI) as a feasible primary imaging modality for diagnosing these conditions [2,3]. Due to its exceptional soft tissue contrast, MRI provides superior diagnostic accuracy over the commonly used contrast-enhanced computed tomography (CECT) [4,5]. This capability allows precise identification and delineation of abscesses and excellent sensitivity in detecting soft tissue edema patterns [6–11].

Previous investigations into acute neck infections have revealed characteristic patterns of reactive edema, especially retropharyngeal (RPE) and mediastinal edema (ME), visualized as hyperintense areas on MRI fluid-sensitive, fat-suppressed T2-weighted sequences [7]. These edema patterns, in conjunction with abscess size, have demonstrated prognostic significance, aiding in predicting intensive care unit admissions and the duration of hospital stay, thus acting as markers of disease severity [7,12]. However, the general applicability of these emerging MRI-based prognostic biomarkers in different clinical settings has yet to be conclusively established. One study reported a reasonable degree of interobserver agreement for these edema patterns between two neuro-radiologists [7]. However, their clinical adoption hinges on whether physicians across specialties can reliably identify them, given the complexity of neck MRI compared to CECT and varying expertise among readers. This question is pivotal for expanding the adoption of these MRI features in routine clinical assessments and enhancing patient management strategies in acute care environments.

To address this question, we initiated a multicenter study with a multidisciplinary, blinded, interobserver design. We aimed to evaluate the accuracy of physicians in identifying RPE and ME using axial T2-weighted Dixon MRI sequences acquired from patients with acute neck infections in the emergency setting. The study involved a diverse group of 28 annotators, including radiology residents, board-certified radiologists, and specialists in otolaryngology and head and neck surgery. By validating the reliability of these imaging biomarkers across diverse expertise levels, we aim to support the integration of MRI into routine practice for acute neck infections.

2. Materials and methods

2.1. Cohort

The study flowchart is shown in Fig. 1. We followed previously published procedures for patient selection, data acquisition, imaging protocol and interpretation, and the extraction of medical and surgical information [6,7,12]. The initial cohort included 535 patients with clinically confirmed acute neck infections who underwent emergency neck MRI between April 1, 2013, and August 30, 2021. Patients were scanned with the Philips Ingenia 3 T system using a dS HeadNeckSpine coil configuration. A gadolinium-based intravenous contrast agent (Dotarem®) was routinely administered. We utilized T1- and T2-weighted images acquired with the Dixon method, diffusion-weighted imaging (DWI), and T1-weighted Dixon sequences after contrast agent administration (see [Supplementary Materials](#) for sequence parameters). Review by the institutional review board or ethics committee was not required because registry-based retrospective studies of existing data are exempt from ethical approval by national law and are only subject to hospital district permission. A waiver for written patient consent was not sought for the same reason. Study permission was obtained from the hospital district board.

From this large cohort of 535 patients, we sampled 60 patients in the test cohort (Table 1). The patients were randomly selected from RPE and ME positive and negative cases to ensure that the prevalence of positive cases for both edema patterns roughly matched those in the total sample. We included axial in-phase and water-only T2-weighted Dixon images

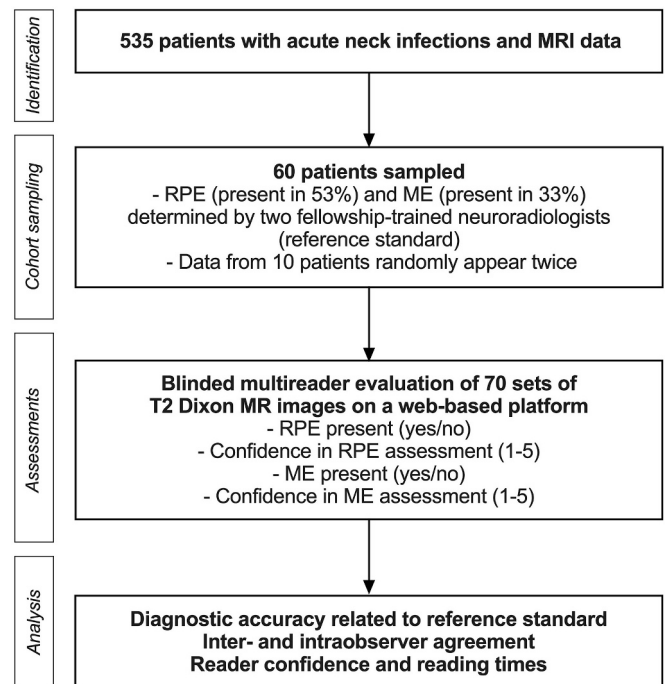


Fig. 1. Study flowchart. Abbreviations: ME, mediastinal edema; RPE, retropharyngeal edema.

Table 1

Patient characteristics.

Characteristics	Patients (N = 60)
Age (years)	39 ± 20
Female (%)	50 %
Pediatric (%)	13 %
Abscess (%)	75 %
CRP (mg/l)	127 ± 100
WBC (x 10 ⁹ /L)	15 ± 7
Length of hospital stay (days)	4 ± 3
RPE (%)	53 %
ME (%)	33 %

Abbreviations: CRP, C-reactive protein; ME, mediastinal edema; N, number; RPE, retropharyngeal edema; WBC, white blood-cell count.

from these 60 patients, with ten randomly selected patients appearing twice in random order, unbeknownst to the readers, to assess intra-observer consistency. The consensus of two fellowship-trained neuro-radiologists with extensive experience and competence in emergency neck MRI, who did not serve as readers, formed the reference standard for RPE and ME. Altogether, 36 of the 60 patients have been previously reported [7]. The previous article addressed the clinical significance of RPE and ME, whereas the intra- and interobserver agreement of RPE and ME are reported in the current study.

2.2. Clinical annotators

A total of 28 clinical annotators (readers) were recruited for the study, including radiology residents, non-neuroradiologist specialists, neuroradiology or head and neck radiology subspecialists, and seven head and neck surgeons (Table 2) from all academic university hospitals in our country. We recorded years of post-residency experience (or years of residency from residents), frequency of emergency MRI use for acute neck infections (1 = seldom, 2 = occasionally, 3 = often), and self-assessed proficiency in interpreting neck MR images (1 = not at all, 2 = somewhat, 3 = good).

Table 2
Reader characteristics.

	Radiology residents	Non-neuro- radiologists	Neuro/HN radiologists	Surgeons	Total
Number	4	5	12	7	28
Years of experience (mean, range)	3.5 (1–5)	13 (11–15)	18 (6–29)	19 (13–30)	15 (1–30)
Outside institute (%)	50 %	20 %	75 %	43 %	54 %
Routine use of emergency neck MRI (mean, SD)	1.0 (1.2)	1.6 (0.9)	0.9 (0.7)	1.4 (0.8)	1.2 (0.8)
Self-evaluated competence (mean, SD)	1.0 (0.0)	1.4 (0.5)	1.6 (0.5)	1.0 (0.0)	1.3 (0.5)
Reading times per case in minutes (mean, range)	2.8 (2.6–3.2)	3.4 (1.6–6.8)	3.2 (1.7–5.5)	2.3 (1.3–3.3)	3.0 (1.3–6.8)

Abbreviations: HN, head and neck; SD, standard deviation.

2.3. Annotation task

From all 70 MR images, the following four ratings were obtained: presence of RPE (yes/no), confidence in the assessment of RPE (1–5, where 1 = very unconfident and 5 = very confident), presence of ME (yes/no), and confidence in the assessment of ME (1–5). The time spent per case was measured from the moment the case was opened to the completion of the annotation, without the readers’ awareness. Readers

were blinded to any clinical, laboratory, or surgical information, as well as to assessments by other readers. The reading task was conducted using a commercially available web-based platform (Collective Minds Radiology), where readers could zoom, adjust windowing, pan, and scroll the images as desired. Before the annotation task, readers were provided with a previous publication on the topic, short written instructions about the edema patterns, and a short tutorial video on using the platform.

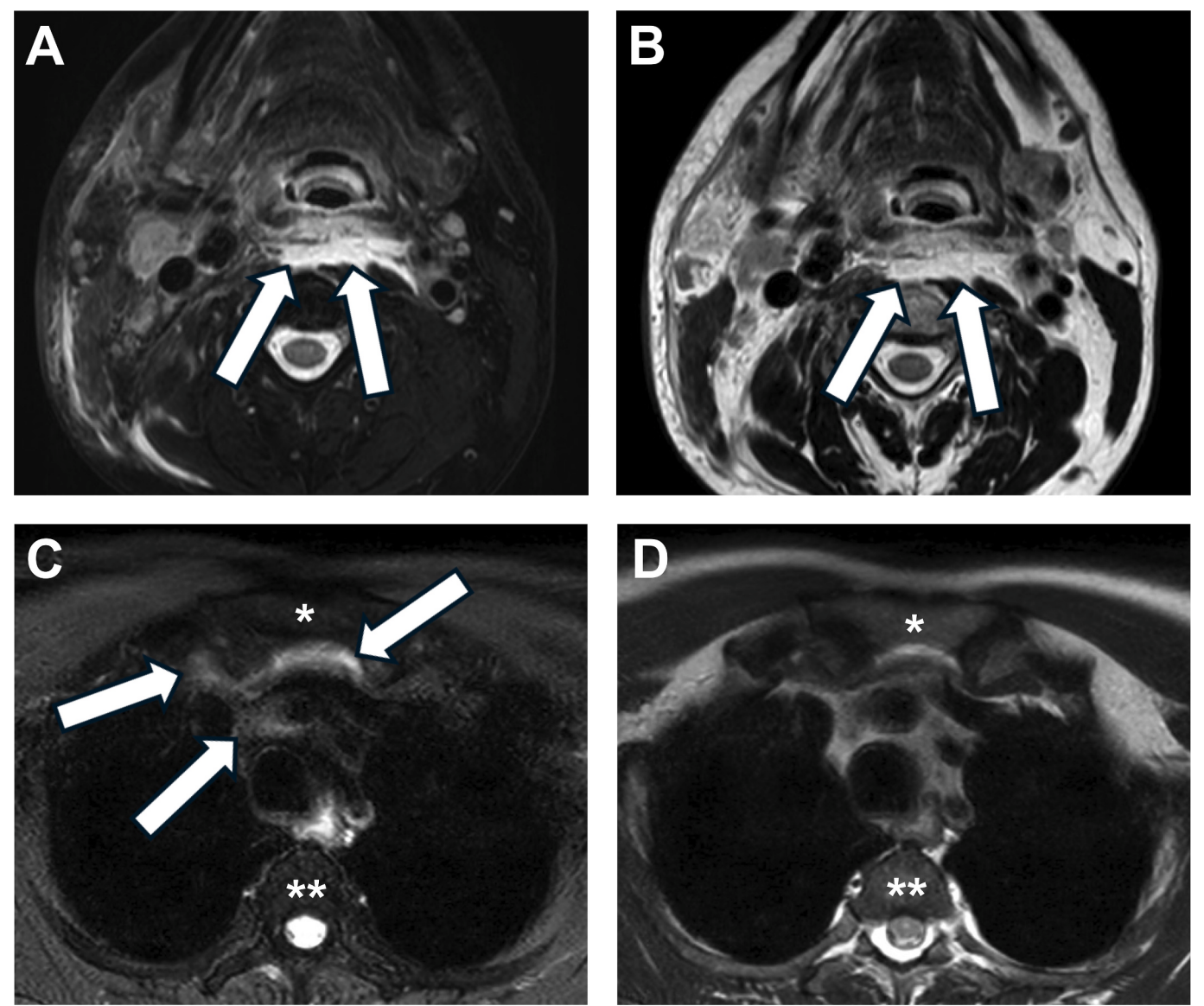


Fig. 2. Examples of RPE (top row) and ME (bottom row). Top row: axial T2 Dixon water-only (A) and in-phase (B) images show RPE as an area of increased T2 signal intensity in the retropharyngeal space (arrows). Bottom row: T2 Dixon water-only (A) and in-phase (B) images at the level of the sternum (*) and the third thoracic vertebra (**) demonstrate ME as an area of increased T2 signal intensity in the upper mediastinum (arrows).

2.4. Definitions of edema patterns

RPE was defined as an area of a hyperintense signal of at least 2 mm in anteroposterior thickness between the prevertebral muscles posteriorly and the superior pharyngeal constrictor muscle anteriorly in at least two consecutive axial fat-suppressed T2-weighted Dixon images [7]. ME was defined as an area of a hyperintense signal in the soft tissues in axial fat-suppressed T2-weighted Dixon images at or below the level of the thoracic inlet, using the superior border of the manubrium sterni or the first thoracic vertebra as the superior border [7]. Examples of RPE and ME are given in Fig. 2.

2.5. Statistical analyses

For all readers, diagnostic accuracy was calculated against the reference standard: sensitivity (true positives / [true positives + false negatives]), specificity (true negatives / [true negatives + false positives]), and accuracy (all true diagnoses / all assessments). These data are expressed as medians and interquartile range (IQR), as normality assumptions could not be met. Confidence values and reading times were treated as normally distributed. Differences in accuracy metrics between reader groups and institutions were assessed using non-parametric Kruskal-Wallis rank sum and Mann-Whitney U-tests. Diagnostic performance was compared across reader groups as well as all radiologists vs. surgeons. Interobserver agreement was evaluated with Fleiss' kappa and compared between groups using the standard error for z-transformations and the normality assumption. Intraobserver consistency (agreement) was assessed using percent accuracy. Linear regression was used to identify associations between confidence and diagnostic metrics, confidence and false diagnoses, and the relationship between time used per case and diagnostic metrics and time and false diagnoses. As reading time was measured without the readers' knowledge, outliers (>90 min, 1 % of timestamps) were identified and excluded. Additionally, time data was excluded from cases opened more than once by the same reader (22.8 % of timestamps). Statistical analyses were performed using IBM SPSS Statistics for Mac (version 28, copyright IBM Corporation 2021) and R version 4.4.0. P-values less than 0.05 were considered statistically significant.

3. Results

The test cohort had a mean age of 39 years (SD 20, range 0–93 years). There was an equal sex distribution with 30 males (50 %) and 30 females (50 %). The cohort included eight pediatric (13 %; <18 years old) and 52 adult (87 %) patients. RPE was present in 32 out of 60 participants (53 %), while ME was present in 20 out of 60 participants (33 %) (Table 1).

For diagnostic performance across all readers, RPE had median sensitivity, specificity, and accuracy of 0.92 (IQR: 0.86–0.95), 0.80 (IQR: 0.73–0.92), and 0.87 (IQR: 0.83–0.90), respectively, whereas ME showed median values of 0.88 (IQR: 0.75–0.96), 0.86 (IQR: 0.71–0.91), and 0.85 (IQR: 0.77–0.91) (Table 3). The median intraobserver consistency was 90.0 % (IQR: 87.5–100.0) for RPE and 90.0 % (IQR: 80.0–100.0) for ME.

In the Kruskal-Wallis test across groups, we found only weak evidence for reader group differences in RPE sensitivity ($p = 0.058$), ME sensitivity ($p = 0.061$), and ME accuracy ($p = 0.072$) and no significant differences in other metrics. Radiologists demonstrated significantly better performance than surgeons in RPE sensitivity (median: 0.92, IQR: 0.89–0.97 vs. 0.81, IQR: 0.81–0.89; $p = 0.007$), RPE accuracy (median: 0.87, IQR: 0.86–0.91 vs. 0.81, IQR: 0.81–0.86; $p = 0.042$), ME sensitivity (median: 0.92, IQR: 0.80–0.96 vs. 0.72, IQR: 0.70–0.80; $p = 0.011$), and ME accuracy (median: 0.86, IQR: 0.80–0.91 vs. 0.74, IQR: 0.69–0.81; $p = 0.024$) (Fig. 3). No significant group differences were seen in intraobserver consistency.

Across all readers, routine use of emergency MRI was not associated

Table 3

Diagnostic accuracy, confidence, and interobserver agreement.

Outcome	Group	Sens.	Spec.	Acc.	Intra	Conf.	Kappa
RPE	Overall	0.92	0.80	0.87	0.90	4.3	0.61
	Rad. resident	0.92	0.85	0.89	0.95	4.1	0.75
	Non-neurorad.	0.92	0.73	0.83	0.90	4.3	0.63
	Neuro/HN-rad	0.93	0.86	0.88	0.95	4.5	0.67
	Surgeon	0.81	0.78	0.81	0.90	4.3	0.47
ME	Overall	0.88	0.86	0.85	0.90	4.1	0.52
	Rad. resident	0.84	0.79	0.81	0.90	3.7	0.51
	Non-neurorad.	0.96	0.87	0.86	0.90	4.2	0.63
	Neuro/HN-rad	0.92	0.90	0.91	0.95	4.3	0.56
	Surgeon	0.72	0.78	0.74	0.80	4.0	0.39

Abbreviations: Acc., accuracy; Conf., confidence; HN, head and neck; Intra, intraobserver consistency; ME, mediastinal edema; rad., radiology/radiologist; RPE, retropharyngeal edema; Sens., sensitivity; Spec., specificity.

with diagnostic accuracy metrics. However, self-evaluated proficiency was positively associated with RPE sensitivity ($p = 0.011$) and ME intraobserver consistency ($p = 0.024$).

Overall, high confidence was found for RPE (mean 4.3) and ME (mean 4.1), with no statistically significant differences in RPE ($p = 0.358$) or ME ($p = 0.082$) across reader groups (Table 3) or between radiologists and non-radiologists for (RPE $p = 0.739$, ME $p = 0.388$).

Fleiss' Kappa indicated substantial interobserver agreement for RPE ($\kappa = 0.61$, 95 % CI [0.54–0.70], $p < 0.001$) and moderate agreement for ME ($\kappa = 0.52$, 95 % CI [0.43–0.61], $p < 0.001$) across all readers (Table 3). Among the groups, radiology residents showed the highest agreement for RPE ($\kappa = 0.75$), while surgeons demonstrated the lowest agreement for both RPE ($\kappa = 0.47$) and ME ($\kappa = 0.39$). Radiologists and neuroradiologists exhibited moderate to substantial agreement for both metrics, with RPE $\kappa = 0.63$ and $\kappa = 0.67$ and ME $\kappa = 0.63$ and $\kappa = 0.56$, respectively. When comparing radiologists (including neuroradiologists) to surgeons, radiologists demonstrated more substantial agreement for both RPE ($\kappa = 0.67$ vs. 0.47) and ME ($\kappa = 0.57$ vs. 0.39, $p < 0.05$ for both). Overall, the agreement was substantial for RPE, moderate for ME, and varied across groups, with radiology residents performing best for RPE and surgeons showing the lowest agreement across both metrics (Table 3).

Linear regression analyses showed a significant positive relationship between mean reader confidence and accuracy ($R^2 = 0.323$, $p = 0.002$) and specificity ($R^2 = 0.26$, $p = 0.005$) for ME. No significant relationships were observed between mean reader confidence and any performance metric for RPE ($p > 0.05$ for all). Additionally, at the case level, confidence and correct answers correlated for both RPE ($R^2 = 0.11$, $p < 0.001$) and ME ($R^2 = 0.11$, $p < 0.001$). Finally, at the case level, cumulative false diagnoses were inversely associated with mean confidence (Fig. 4).

The average reading time was 3 min per case, without significant group differences ($p = 0.412$ for group comparison, $p = 0.116$ for radiologists vs. surgeons) (Table 2). A linear regression analysis was performed to evaluate the relationship between time spent per case (in minutes) and error (false diagnosis) count (no errors, one error [RPE or ME], or two errors [RPE and ME]). The model revealed a small but statistically significant positive association between time and error count ($R^2 = 0.02$, $p < 0.001$). For every additional minute spent on a case, the error count increased by approximately 0.018 errors. Linear regression also showed a significant decrease in time spent per case as a function of case number, starting at 4 min 17 s (257 s) for the first case and decreasing by 2.2 s per case ($R^2 = 0.24$, $p < 0.001$).

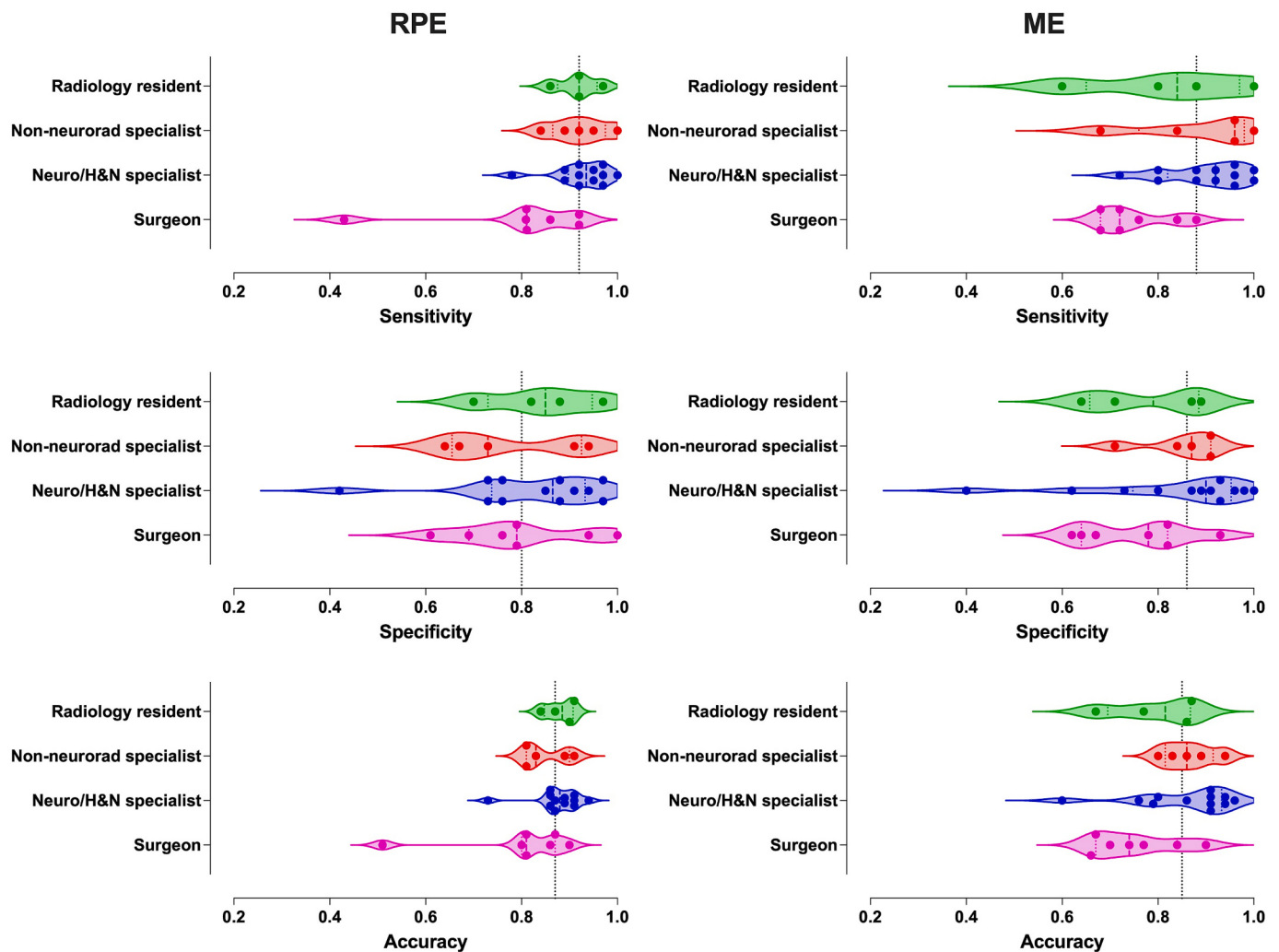


Fig. 3. Violin plots of sensitivity, specificity, and accuracy of RPE (left column) and ME (right column) with individual values (closed circles) in different reader groups. Inside the violins, dashed lines are group medians, and dotted lines are group 25 % and 75 % quartiles. Dotted lines behind the violins represent overall medians across all readers. Abbreviations: HN, head and neck; ME, mediastinal edema; rad., radiology/radiologist; RPE, retropharyngeal edema.

4. Discussion

MRI is an emerging tool for acute neck infections, revealing specific edema patterns with clinical and prognostic significance. In this study, 28 readers across specialties achieved high diagnostic accuracy (0.87 for RPE, 0.85 for ME) and substantial interobserver agreement ($\kappa = 0.61$ for RPE, 0.52 for ME) on T2-weighted Dixon sequences. These findings suggest that emerging MRI biomarkers from complex anatomical imaging can be reliably applied across different levels of expertise, institutions, and medical disciplines, aiding risk stratification and treatment planning.

The reliable detection of RPE and ME could be integrated into clinical pathways for acute neck infections. For example, patients with confirmed RPE or ME on MRI could be prioritized for early surgical consultation or closer monitoring, given their association with severe disease [7,12]. Incorporating these MRI biomarkers into triage protocols could optimize resource allocation and improve patient outcomes in emergency settings.

Retropharyngeal edema (RPE) and mediastinal edema (ME) are distinct patterns of edema that show prognostic significance in patients with acute neck infections, indicative of the severity of the disease; these edema patterns are associated with patient-centered outcomes, such as ICU admission and prolonged hospital stay [7]. While this study focuses on diagnostic accuracy and agreement, these prior findings underscore

the clinical relevance of reliable edema detection. Although MRI is superior to CECT in detecting drainable neck abscesses [13,14], the diagnostic performance of CECT for RPE and ME is currently unknown. This is an important implication for future studies, given that CT is the most common emergency modality for acute neck infections [3].

Specificity tended to be lower than sensitivity, especially for RPE. High sensitivity suggests a good ability to detect these edema patterns and alert about the possibility of a severe course of illness. Lower specificity indicates the potential for false positives. ME demonstrated lower agreement between readers and higher variability in diagnostic accuracy compared to RPE. Detecting edema in the upper mediastinum from neck scans is inherently challenging due to its variable presentation and less distinct anatomical boundaries. Technical limitations include suboptimal fat saturation or limited upper mediastinal visualization. A previous study in a single institution found that in 16 % of patients, the images did not sufficiently cover the upper mediastinum to evaluate ME [7]. Radiologists outperformed surgeons in agreement and diagnostic accuracy, reflecting expertise in complex imaging, though all groups showed strong performance. Notably, radiology residents matched specialists, suggesting that brief training (e.g., our tutorial video) enables non-experts to interpret these patterns effectively, a key step toward broader MRI use.

Our diagnostic accuracy and interobserver agreement findings are consistent with those published previously regarding MR imaging

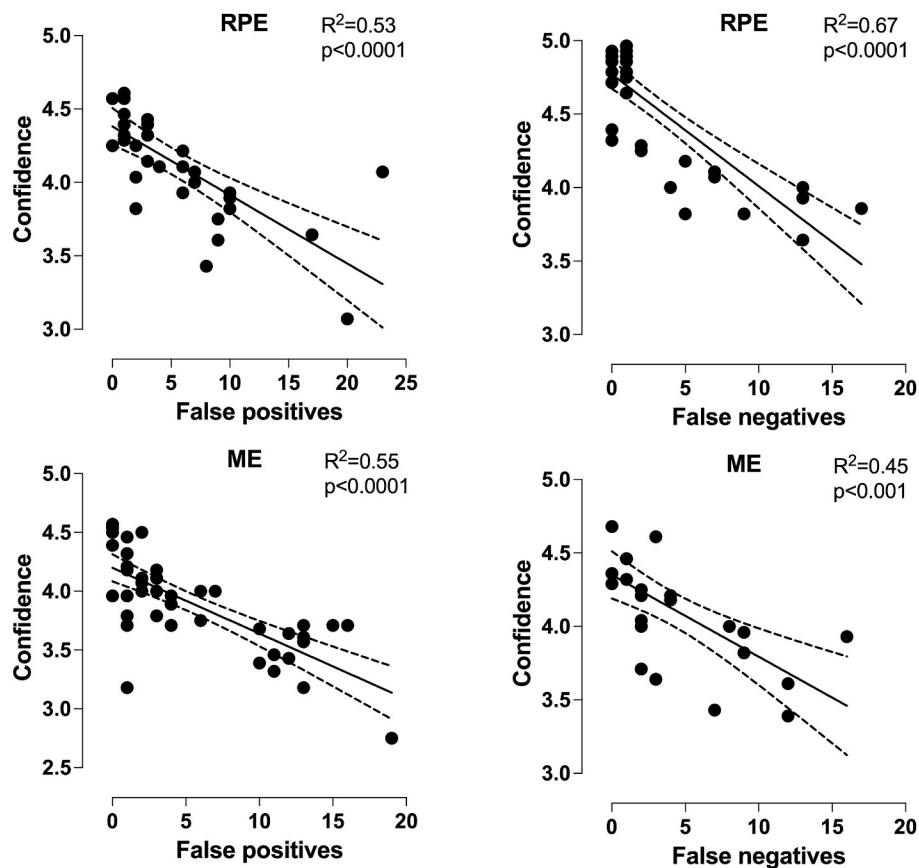


Fig. 4. Inverse correlations between cumulative false diagnoses and mean confidence scores at the case level.

outcomes. Prior studies have demonstrated that accuracy and agreement can improve significantly with more MRI expertise [15,16], while others have highlighted substantial variability in detection performance, even among experienced radiologists [17]. Systematically assessing the variability of human performance metrics is essential, especially for novel imaging biomarkers. Few interobserver studies have focused on the neck soft tissues. One study found moderate agreement among 10 experienced head and neck radiologists in detecting recurrent head and neck cancer [18]. Radiologist expertise improves the detection of bone invasion from oral cancer [19] and the classification of salivary gland tumors [20]. Our results are consistent with previous studies showing good agreement between readers in detecting soft tissue and bone edema on fat-suppressed T2-weighted images [21–24]. As expected, diagnostic accuracy correlated positively with self-assessed proficiency, confidence, and short reading times.

This study utilized a sizable dataset of MRI scans corroborated by clinical assessments and surgical outcomes. While typical multireader studies rely on a handful of radiologists, we incorporated a broad spectrum of annotators, including radiologists of varying experience levels, residents, and specialists from related fields. A web-based platform facilitated a large-scale, multi-institutional study by providing an optimal infrastructure for managing the annotation process, data collection, and analysis.

A primary limitation of this research is the use of a subjective reference standard for RPE and ME. While previous work has shown correlations with clinical outcomes and substantial agreement among expert readers [7], the subjectivity in defining these markers remains a concern. This limitation is shared by many semi-quantitative imaging biomarkers [25] and does not affect inter- or intraobserver agreement. Quantitative thresholds could refine this in future work. The absence of histopathologic or surgical correlation for RPE and ME is a limitation, as these edema patterns are primarily imaging-based and not confirmed

surgically. Prior work [7,12] has validated their prognostic significance for ICU admission and hospital stay length, supporting their clinical relevance. In our study, detecting edema patterns was not dependent on radiologists' expertise, as residents performed as well as specialists. We focused exclusively on T2-weighted axial sequences, sufficient for detecting RPE and ME. However, additional imaging planes and sequences might improve accuracy in specific clinical scenarios, such as when an enlarged retropharyngeal lymph node mimics RPE. MRI data were acquired using a single Philips Ingenia 3 T scanner with T2-weighted Dixon sequences, limiting generalizability to other MRI vendors, field strengths, or sequence protocols. Variations in sequence parameters and image quality, such as those affecting fat suppression quality, may affect the conspicuity of RPE and ME, particularly in the upper mediastinum, where suboptimal fat suppression or limited anatomical coverage can reduce diagnostic performance. Future studies should validate the detection of RPE and ME across multicenter imaging datasets with diverse MRI platforms and sequence configurations to confirm feasibility and robustness in varied clinical settings. Time data were collected retrospectively, with readers unaware that their time was being measured. Some outliers were observed, possibly because readers did not account for time in their evaluations. Consequently, the time data should be regarded as approximate. A practical limitation is the low availability of MRI as the primary imaging tool in emergencies, hindering the widespread adoption of this approach in imaging acute neck infections. The absence of a formal power calculation is a limitation, and the sample size may be underpowered to detect subtle differences between reader groups, particularly for ME, which has lower prevalence. Future studies with larger, powered samples are needed to validate these results. The test cohort of 60 patients was non-consecutively sampled from a registry of 535 patients to approximate the population prevalence of RPE (52 %) and ME (24 %), as reported previously [12]. The test set prevalences (53 % for RPE, 33 % for ME) closely align for RPE but

slightly overestimate ME, potentially introducing minor selection bias.

In conclusion, this multicenter study demonstrates high diagnostic accuracy and substantial interobserver agreement in detecting MRI-based edema patterns (RPE and ME) in acute neck infections across a diverse group of radiologists and surgeons. Radiologists achieved higher sensitivity and agreement compared to surgeons, reflecting their expertise in complex imaging interpretation. These findings validate MRI as a reliable tool for risk assessment, particularly for radiologists, supporting its integration into emergency imaging protocols where accessible. Further training and validation may be needed to improve surgeon performance for broader clinical application.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (GPT-4 architecture), Grok 2 beta, and Grammarly Pro in order to enhance the grammatical accuracy and semantic clarity of portions of the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

CRediT authorship contribution statement

Jari-Pekka Vierula: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Katri Aro:** Writing – review & editing, Data curation. **Michaela Bode:** Writing – review & editing, Data curation. **Eeva Haapio:** Writing – review & editing, Data curation. **Juhana Hakumäki:** Writing – review & editing, Data curation. **Tatu Happonen:** Writing – review & editing, Data curation. **Carita Hollmén:** Writing – review & editing, Data curation. **Heidi Huhtanen:** Writing – review & editing, Data curation. **Heikki Irjala:** Writing – review & editing, Data curation. **Anna Jouppila-Mättö:** Writing – review & editing, Data curation. **Ville Jussila:** Writing – review & editing, Data curation. **Kalle Kaapu:** Writing – review & editing, Data curation. **Jari O. Karhu:** Writing – review & editing, Data curation. **Petri Koivunen:** Writing – review & editing, Data curation. **Marcin Kolasa:** Writing – review & editing, Data curation. **Antti Kovalainen:** Writing – review & editing, Data curation. **Goran Kurdo:** Writing – review & editing, Data curation. **Eero Kytö:** Writing – review & editing, Data curation. **Tiina Laurikainen:** Writing – review & editing, Data curation. **Riikka Lindén:** Writing – review & editing, Data curation. **Ilkka Martikainen:** Writing – review & editing, Data curation. **Kristofer Nyman:** Writing – review & editing, Data curation. **Timo Paananen:** Writing – review & editing, Data curation. **Teemu Paavilainen:** Writing – review & editing, Data curation. **Jarkko Saravo:** Writing – review & editing, Data curation. **Jutta Tähtinen:** Writing – review & editing, Data curation. **Ville Tammilehto:** Writing – review & editing, Data curation. **Jarno Velhonoja:** Writing – review & editing, Data curation. **Niklas Ånäs:** Writing – review & editing, Data curation. **Aapo Sirén:** Writing – review & editing, Data curation. **Jaakko Heikkinen:** Writing – review & editing, Data curation. **Janne Nurminen:** Writing – review & editing, Data curation, Conceptualization. **Mikko Nyman:** Writing – review & editing, Conceptualization. **Kimmo Mattila:** Writing – review & editing, Conceptualization. **Jussi Hirvonen:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Jussi Hirvonen reports financial support was provided by Sigrid Jusélius Foundation. Given their role as Section Editor, Jussi Hirvonen had no involvement in the peer-review of this article and has no access to

information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they 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 A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejrad.2025.112526>.

References

- [1] F. Vieira, S.M. Allen, R.M.S. Stocks, J.W. Thompson, Deep neck infection, *Otolaryngol. Clin. North Am.* 41 (2008) 459–483, <https://doi.org/10.1016/j.otc.2008.01.002>.
- [2] J. Hirvonen, J. Heikkinen, M. Nyman, T. Happonen, J. Velhonoja, H. Irjala, T. Soukka, K. Mattila, J. Nurminen, MRI of acute neck infections: evidence summary and pictorial review, *Insights Imag.* 14 (2023) 5, <https://doi.org/10.1186/s13244-022-01347-9>.
- [3] J. Hirvonen, R.K. Lingam, S. Connor, E.S.R. Essentials, acute infections of the head and neck—practice recommendations by the European Society of Head and Neck Radiology, *Eur. Radiol.* (2025), <https://doi.org/10.1007/s00330-025-11818-4>.
- [4] J. Hagelberg, B. Pape, J. Heikkinen, J. Nurminen, K. Mattila, J. Hirvonen, Diagnostic accuracy of contrast-enhanced CT for neck abscesses: a systematic review and meta-analysis of positive predictive value, *PLoS One* 17 (2022) e0276544, <https://doi.org/10.1371/journal.pone.0276544>.
- [5] S. Chuang, H. Lin, Y. Wen, F. Hsu, Pitfalls of CT for deep neck abscess imaging assessment: a retrospective review of 162 cases, *B-ENT* 9 (2013) 45–52.
- [6] J. Nurminen, J. Velhonoja, J. Heikkinen, T. Happonen, M. Nyman, H. Irjala, T. Soukka, K. Mattila, J. Hirvonen, Emergency neck MRI: feasibility and diagnostic accuracy in cases of neck infection, *Acta Radiol.* 62 (2021) 735–742, <https://doi.org/10.1177/0284185120940242>.
- [7] J. Heikkinen, J. Nurminen, J. Velhonoja, H. Irjala, T. Happonen, T. Soukka, K. Mattila, J. Hirvonen, Clinical and prognostic significance of emergency MRI findings in neck infections, *Eur. Radiol.* 32 (2022) 1078–1086.
- [8] J. Heikkinen, J. Nurminen, J. Velhonoja, H. Irjala, T. Soukka, T. Happonen, M. Nyman, K. Mattila, J. Hirvonen, MRI findings in acute tonsillar infections, *Am. J. Neuroradiol.* 43 (2022) 286–291, <https://doi.org/10.3174/ajnr.A7368>.
- [9] J. Heikkinen, V. Jokihaka, J. Nurminen, V. Jussila, J. Velhonoja, H. Irjala, T. Soukka, T. Happonen, J. Järnstedt, M. Nyman, K. Mattila, J. Hirvonen, MRI of odontogenic maxillofacial infections: diagnostic accuracy and reliability, *Oral Radiol.* 39 (2023) 364–371, <https://doi.org/10.1007/s11282-022-00646-7>.
- [10] J. Nurminen, J. Heikkinen, T. Happonen, J. Velhonoja, H. Irjala, T. Soukka, L. Ivaska, K. Mattila, J. Hirvonen, Magnetic resonance imaging findings in pediatric neck infections—a comparison with adult patients, *Pediatr. Radiol.* 52 (2022) 1158–1166, <https://doi.org/10.1007/s00247-021-05275-6>.
- [11] J.-P. Vierula, J. Nurminen, V. Jussila, M. Nyman, J. Heikkinen, B. Pape, K. Sorvettula, K. Mattila, J. Hirvonen, Diagnostic performance of short noncontrast biparametric 3-T MRI for tonsillar infections: comparison with a full protocol including contrast-enhanced sequences, *Eur. Radiol. Exp.* 7 (2023) 65, <https://doi.org/10.1186/s41747-023-00379-0>.
- [12] J.-P. Vierula, H. Merisaari, J. Heikkinen, T. Happonen, A. Sirén, J. Velhonoja, H. Irjala, T. Soukka, K. Mattila, M. Nyman, J. Nurminen, J. Hirvonen, MRI-based risk factors for intensive care unit admissions in acute neck infections, *Eur. J. Radiol. Open* 14 (2025) 100648, <https://doi.org/10.1016/j.ejro.2025.100648>.
- [13] A. Muñoz, M. Castillo, M.A. Melchor, R. Gutiérrez, Acute neck infections: prospective comparison between CT and MRI in 47 patients, *J. Comput. Assist. Tomogr.* 25 (2001) 733–741, <https://doi.org/10.1097/00004728-200109000-00011>.
- [14] V.R. Babu, S. Ikkurthi, D.K. Perisetty, K.A.S. Babu, M. Rasool, S. Shaik, A prospective comparison of computed tomography and magnetic resonance imaging as a diagnostic tool for maxillofacial space infections, *J. Int. Soc. Prev. Community Dent.* 8 (2018) 343–348, <https://doi.org/10.4103/jispcd.JISPCD.315.17>.
- [15] M.M.N. Leeuwenburgh, B.M. Wiarda, S. Jensch, H.W. Van Es, H.B.A.C. Stockmann, J.W.C. Gratama, L.P.J. Cobben, P.M.M. Bossuyt, M.A. Boermeester, J. Stoker, Accuracy and interobserver agreement between MR-non-expert radiologists and MR-experts in reading MRI for suspected appendicitis, *Eur. J. Radiol.* 83 (2014) 103–110, <https://doi.org/10.1016/j.ejrad.2013.09.022>.
- [16] K.H. Lee, H.J. Park, S.Y. Lee, M.S. Kim, Y.T. Lee, Y.B. Kim, J.N. Kim, I. Yoon, Interobserver reliability and clinical validity of the MRI grading system for cervical central stenosis based on sagittal T2-weighted image, *Eur. J. Radiol.* 127 (2020) 108987, <https://doi.org/10.1016/j.ejrad.2020.108987>.
- [17] J.H. Min, M.W. Lee, H.S. Park, D.H. Lee, H.J. Park, S. Lim, S.-Y. Choi, J. Lee, J. E. Lee, S.Y. Ha, D.I. Cha, K.C. Carriere, J.H. Ahn, Interobserver variability and diagnostic performance of gadoteric Acid-enhanced MRI for predicting microvascular invasion in hepatocellular carcinoma, *Radiology* 297 (2020) 573–581, <https://doi.org/10.1148/radiol.2020201940>.
- [18] B. Peltenburg, J.P. Driessen, M.E.P. Philipps, F.A. Pameijer, J. Castelijns, J. W. Dankbaar, B. Dorgelo, C.A.H. Lange, F.J.A. Meijer, W.M. Palm, S.D. Roosendaal, S.C.A. Steens, B.M. Verbit, I. Stegeman, L.M. Janssen, R. De Bree, C.H.J. Terhaard,

- The interobserver agreement in the detection of recurrent HNSCC using MRI including diffusion weighted MRI, *Eur. J. Radiol.* 105 (2018) 134–140, <https://doi.org/10.1016/j.ejrad.2018.05.031>.
- [19] M. Silva, E.I. Zambrini, G. Chiari, I. Montermini, C. Manna, T. Poli, D. Lanfranco, E. Sesenna, E. Thai, N. Sverzellati, Pre-surgical assessment of mandibular bone invasion from oral cancer: comparison between different imaging techniques and relevance of radiologist expertise, *Radiol. Med. (Torino)* 121 (2016) 704–710, <https://doi.org/10.1007/s11547-016-0654-1>.
- [20] T.J. Vogl, M.H. Albrecht, N.-E.-A. Nour-Eldin, H. Ackermann, A. Maataoui, T. Stöver, M.W. Bickford, T. Stark-Paulsen, Assessment of salivary gland tumors using MRI and CT: impact of experience on diagnostic accuracy, *Radiol. Med. (Torino)* 123 (2018) 105–116, <https://doi.org/10.1007/s11547-017-0813-z>.
- [21] M.H. Dagestad, M.K. Toppe, P.M. Kristoffersen, E. Gjefsen, E. Andersen, J. Assmus, N. Vetti, A. Espeland, Dixon T2 imaging of vertebral bone edema: reliability and comparison with short tau inversion recovery, *Acta Radiol.* 65 (2024) 273–283, <https://doi.org/10.1177/02841851221146130>.
- [22] P. Mehta, R. Patel, R. Bhadelia, Y.-M. Chang, A. Brook, C.-W.-C. Huang, K. Manzoor, L. Hacin-Bey, V. Ivanovic, Paraspinal soft tissue edema ratio: an accurate marker for early lumbar spine spondylodiscitis on an unenhanced MRI, *Clin. Imaging* 86 (2022) 38–42, <https://doi.org/10.1016/j.clinimag.2022.03.009>.
- [23] P.M. Kristoffersen, N. Vetti, K. Storheim, L.C. Bråten, M.P. Rolfsen, J. Assmus, A. Espeland, Short tau inversion recovery MRI of Modic changes: a reliability study, *Acta Radiol. Open* 9 (2020) 2058460120902402, <https://doi.org/10.1177/2058460120902402>.
- [24] V.F. Dalto, R.L. Assad, M.D. Crema, P. Louzada-Junior, M.H. Nogueira-Barbosa, MRI assessment of bone marrow oedema in the sacroiliac joints of patients with spondyloarthritis: is the SPAIR T2w technique comparable to STIR? *Eur. Radiol.* 27 (2017) 3669–3676, <https://doi.org/10.1007/s00330-017-4746-7>.
- [25] European Society of Radiology, N.M. deSouza, E. Achten, A. Alberich-Bayarri, F. Bamberg, R. Boellaard, O. Clément, L. Fournier, F. Gallagher, X. Golay, C. P. Heussel, E.F. Jackson, R. Manniesing, M.E. Mayerhofer, E. Neri, J. O'Connor, K. K. Oguz, A. Persson, M. Smits, E.J.R. Van Beek, C.J. Zech, Validated imaging biomarkers as decision-making tools in clinical trials and routine practice: current status and recommendations from the EIBALL* subcommittee of the European Society of Radiology (ESR), *Insights Imaging* 10 (2019) 87, <https://doi.org/10.1186/s13244-019-0764-0>.