I J Radiol. 2026 March; 23(1): e168277 https://doi.org/10.5812/iranjradiol-168277

Published Online: 2026 May 3 **Systematic Review**

Comparative Image Quality of Deep Learning Reconstruction Against Conventional Reconstruction for T2-weighted Prostate Imaging: A Systematic Review and Meta-Analysis

Narges Azizi [1], Hoda Borooghani [2], Manouchehr Nasrollahzadeh Saravi [3], Sina Delazar [1,*]

[1] Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Tehran University of Medical Sciences, Tehran, Iran
[2] Iran University of Medical Science, Tehran, Iran
[3] Shahid Beheshti University of Medical Science, Tehran, Iran

[*] ***Corresponding Author:*** Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Tehran University of Medical Sciences, Tehran, Iran. Email: sina.delazar.md@gmail.com

Received: 22 November, 2025; **Revised:** 26 February, 2026; **Accepted:** 6 April, 2026

Abstract

Context: Deep learning reconstruction is increasingly proposed to accelerate prostate T2-weighted magnetic resonance imaging (MRI) acquisition without loss of diagnostic quality. However, the magnitude of image quality benefits and the influence of deep learning network design remain unclear.

Objectives: We aimed to compare the image quality of deep learning versus conventional prostate T2 MRI reconstruction and assess how acceleration factors and network architecture influence performance.

Methods: A PRISMA search of PubMed/Medline, Embase, Web of Science, and Scopus (22 February 2026) identified studies comparing deep learning with conventional T2 MRI reconstruction. Paired standardized mean change (SMCC) values were pooled with a three-level random-effects model from extracted paired image-quality scores; meta-regressions examined scan-time acceleration. Small-study bias was tested with a precision-effect method.

Results: Five studies (602 participants; 20 comparisons) included in the study. Overall deep learning reconstruction benefit was +0.14 standard deviation (SD) [95% confidence interval (CI) -2.00 to 2.28; I^2 = 99 %]. At equal scan times deep learning reconstruction improved quality by +2.09 SD; for each fold increase in acceleration, image quality declined by 0.48 SD (95% CI -0.91 to -0.05; P = 0.030). C-SENSE AI at acceleration factors of 1.7, 3.4, and 4.8 demonstrated superior performance, whereas CycleGAN showed inferior results. Small-study effects were evident.

Conclusion: Deep learning reconstruction can increase prostate T2-weighted MRI speed without significant changes in overall image quality; however, overly aggressive implementations may lead to image degradation. Careful validation and protocol-specific tuning are therefore essential prior to clinical adoption.

Keywords: prostate, MRI, deep learning, T2 imaging, image reconstruction, meta-analysis

1. Context

Prostate cancer is one of the most frequent cancers among men and has high mortality and morbidity rates globally (1-3). Although many cases show slow progression, some forms are more aggressive, necessitating precise risk assessment and treatment (1, 4-5). Magnetic resonance imaging (MRI) now facilitates prostate cancer diagnosis, localization, staging, and follow-up because of its suitable soft-tissue contrast; multiparametric MRI (mpMRI) remains the standard protocol, whereas biparametric MRI (bpMRI) provides a faster, contrast-free alternative with fewer side effects (4-8).

T2-weighted imaging (T2WI) is part of both mpMRI and bpMRI and provides important anatomical details of prostate zones (9). In daily practice, image quality fluctuates because sequence parameters, patient motion, and noise can blur fine details. Although conventional reconstruction methods, such as fast spin-

How to Cite: Azizi N, Borooghani H, Nasrollahzadeh Saravi M, Delazar S. Comparative Image Quality of Deep Learning Reconstruction Against Conventional Reconstruction for T2-weighted Prostate Imaging: A Systematic Review and Meta-Analysis. I J Radiol. 2026;23(1):e168277. doi: https://doi.org/10.5812/iranjradiol-168277

echo reconstruction with parallel imaging or compressed sensing, partially recover lost signal, residual noise and artifacts may still obscure small tumors (6, 10-11).

Recent advances in deep learning have changed MRI reconstruction by improving the signal-to-noise ratio (SNR), spatial resolution, and acquisition time (3, 6, 9, 12). Nevertheless, current deep learning implementations still entail trade-offs: performance may vary with deep learning architecture, parameter changes, or scanners, and the balance between accelerated acquisition and preserved image quality remains incompletely defined (12-14).

2. Objectives

We conducted a systematic review and meta-analysis comparing deep learning versus conventional T2-weighted reconstructions in prostate MRI, with the aim of determining whether deep learning techniques provide clinically meaningful gains in image quality and pinpointing the architectures most suitable for routine practice.

3. Methods

The study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (15-16) and was registered in the PROSPERO database (CRD42024613297).

3.1. Search Strategy

A systematic literature search was conducted across PubMed/MEDLINE, Embase, Web of Science, and Scopus for studies published up to February 22, 2026. We included original English-language articles using the following MeSH terms: 'Prostate' OR 'Prostatic Neoplasms'; 'Magnetic Resonance Imaging' OR 'Diagnostic Imaging'; and 'Deep Learning' OR 'Artificial Intelligence' OR 'Image Processing, Computer-Assisted' OR 'Image Enhancement' (see Supplementary Information, Tables S1, S2, and S3). In addition, the reference lists and citations of relevant articles were manually screened.

3.2. Study Selection

Two reviewers independently screened all retrieved studies, with any disagreements resolved by a third reviewer. We included original, English-language studies, in which men underwent prostate imaging and assessed deep learning reconstruction of T2-weighted MRI compared to conventional reconstruction methods. Studies were eligible only if they reported image quality assessment as an outcome by mean and standard deviation (SD) for both of the deep learning model and the conventional method. Reviews, conference proceedings, non-English studies, and studies lacking the required data were excluded.

3.3. Data Extraction and Study-Level Internal Validity

Data extraction was performed independently by two blinded reviewers using a standardized spreadsheet template, with discrepancies resolved by a third reviewer. Extracted data included study characteristics, MRI acquisition parameters (both conventional and deep learning reconstructions), sequence types, reconstruction models, image quality metrics (mean and SD for image noise, sharpness, lesion detectability, artifacts, overall image quality, and diagnostic confidence), and other technical parameters such as voxel size, acquisition time, field strength, and bandwidth. The complete data structure is available in (Online spreadsheet). Study-level internal validity was assessed using the QUADAS tool by two independent reviewers and visualized using the QUADAS module in Stata 17 (StataCorp, College Station, TX, USA) (17).

3.4. Statistical Analysis

Paired Standardized Mean Change: Image quality was treated as a pseudo-continuous variable. The within-group standardized mean change (SMCC) served as the effect size, accounting for the paired nature of observations and the varying number of levels across Likert scales. The correlation between paired observations, required for standard error estimation, was empirically determined as $r = 0.85$ using in-house data. A three-level random-effects meta-analysis was performed to account for multiple, non-independent observations within the same cohort. All analyses were implemented using the metafor package (version 4.8-0) in R (version 4.4.1). The heterogeneity was quantified using the I^2 statistic.

3.5. Leave-One-Out Analysis

Leave-One-Out Analysis: A leave-one-study-out sensitivity analysis was performed to assess the robustness of the pooled effect estimates. The three-level random-effects meta-analysis was re-estimated repeatedly after sequential exclusion of each study, while preserving the original hierarchical structure with random effects specified at the study, model-within-study, and reader-within-study levels. Changes in the pooled SMCC, 95% confidence intervals, and statistical significance were examined across all

iterations to identify potential undue influence of individual studies.

3.6. Meta-Regression by Acceleration

Meta-Regression by Acceleration: To evaluate whether the degree of scan acceleration influenced image quality, we performed a meta-regression. The acceleration factor was included as a continuous moderator within the three-level random-effects analysis. The intercept represents the pooled SMCC when acquisition speed is equivalent to the reference protocol (acceleration = 1), and the slope quantifies the change in SMCC associated with each one-unit increase in acceleration. The omnibus QM test was used to assess whether acceleration explained a significant portion of between-effect heterogeneity. Statistical significance was defined as $P \leq 0.05$, and meta-regression results are illustrated using bubble plots.

3.7. Risk of Bias Assessment

Risk of Bias Assessment: Small-study effects and publication bias were examined with a multilevel funnel plot and a Precision-Effect Test (PET). The PET intercept gives the bias-adjusted SMD at infinite precision, and a positive slope with $P \leq 0.05$ suggests selective reporting due to the larger effects in less precise studies. Risk of bias is illustrated with a funnel plot overlaid by the Precision-Effect Test regression line.

4. Results

The meta-analysis includes twenty paired deep learning versus conventional comparisons from five studies, comprising 602 participants, with a mean age of 66.6 ± 1.74 years (Figure 1). Patients with suspected prostate cancer, histologically proven prostate cancer, and screening or follow-up imaging were included for prostate imaging. The regional distribution of studies consists of three from the United States, one from Germany, and one from the United Kingdom. Patient characteristics are summarized in Table 1. Study-level internal validity shows most concerns centered on reporting indeterminate results, whereas reference and test-execution details were sound in nearly all papers. The distribution of ratings is illustrated in Figures S1 and S2.

4.1. MRI Sequences and Deep Learning Methods

The included studies evaluated deep learning-based reconstruction methods primarily in T2-weighted prostate MRI and axial planes (Table 2). Conventional turbo/fast spin-echo (TSE/FSE) T2-weighted sequences served as the reference standard across all studies. Deep learning approaches encompassed a range of architectures and vendor-specific implementations, including variational networks (Siemens), GAN-based super-resolution models (CycleGAN), compressed sensing combined with AI (C-SENSE AI) at 1.7, 3.4, and 4.8 acceleration factors, and AIR Recon DL (GE Healthcare) applied with low, medium, and high denoising levels and acquisition of standard, fast, and high-resolution. All examinations were performed at 3-Tesla field strength on systems from Siemens, Philips, or GE Healthcare.

4.2. Overall Image-Quality Paired Standardized Mean Change

Overall Image-Quality Paired Standardized Mean Change: The pooled SMCC for overall image quality was 0.14, which did not reach statistical significance [SE = 1.11; z = 0.15; P = 0.88; 95% confidence interval (CI) -2.00 to 2.28] (Figure 2). Heterogeneity was substantial (Q_{19} = 854.3, $P < 0.001$). Variance component analysis demonstrated that the largest proportion of true variability arose from between-study differences (σ^2 = 1.91), followed by variability between models within studies (σ^2 = 0.31) and readers within studies (σ^2 = 0.25). When expressed relative to total variance, heterogeneity was almost entirely attributable to real between-unit differences, yielding a total I^2 of 99.9%. Sensitivity analyses assuming alternative within-pair correlations (r = 0.70 and 0.50) produced comparable pooled estimates, indicating robustness of the findings. Sensitivity analyses assuming lower within-pair correlations yielded comparable results. Using r = 0.70, the pooled SMCC was 0.12 (SE 0.80; z = 0.15, P = 0.88; 95% CI -1.44 to 1.68), and using r = 0.50, the pooled SMCC was 0.10 (SE 0.63; z = 0.16, P = 0.87; 95% CI -1.12 to 1.33). In all sensitivity analyses, the overall effect remained statistically non-significant, with overlapping confidence intervals and substantial residual heterogeneity.

4.3. Leave-One-Out Analysis

Leave-One-Out Analysis: Leave-one-study-out sensitivity analysis at the study level showed that the overall results were robust to the exclusion of any single study. After sequentially removing each study, the pooled SMCC estimates ranged from -0.42 to 0.53, and in all iterations the 95% confidence intervals crossed zero (all p-values > 0.30). No individual study exerted a disproportionate influence on the pooled effect estimate (Supplementary figure S3).

Table 1. Demographic and Clinical Characteristics of Patients in the Included Studies

Study	Year	Country	Participants, No.	Age (mean ± SD)	PSA (ng/mL)	Included Patients
Liu et al. (22)	2021	United States	346	N/A	N/A	Suspected prostate cancer
Johnson et al. (23)	2022	United States	113	68.0 ± 7.0	N/A	Suspected prostate cancer
Harder et al. (21)	2022	Germany	23	64.4 ± 6.2	14.4 ± 18.4	Histologically proven prostate cancer
Tong et al. (18)	2023	United States	80	66 ± N/A	N/A	Suspected prostate cancer
Lee et al. (25)	2023	United Kingdom	40	66 ± N/A	4.70 (median; IQR 2.9)	Suspected prostate cancer

Abbreviation: N/A, Not Available; PSA, prostate specific antigen; SD, standard deviation; IQR, interquartile range.

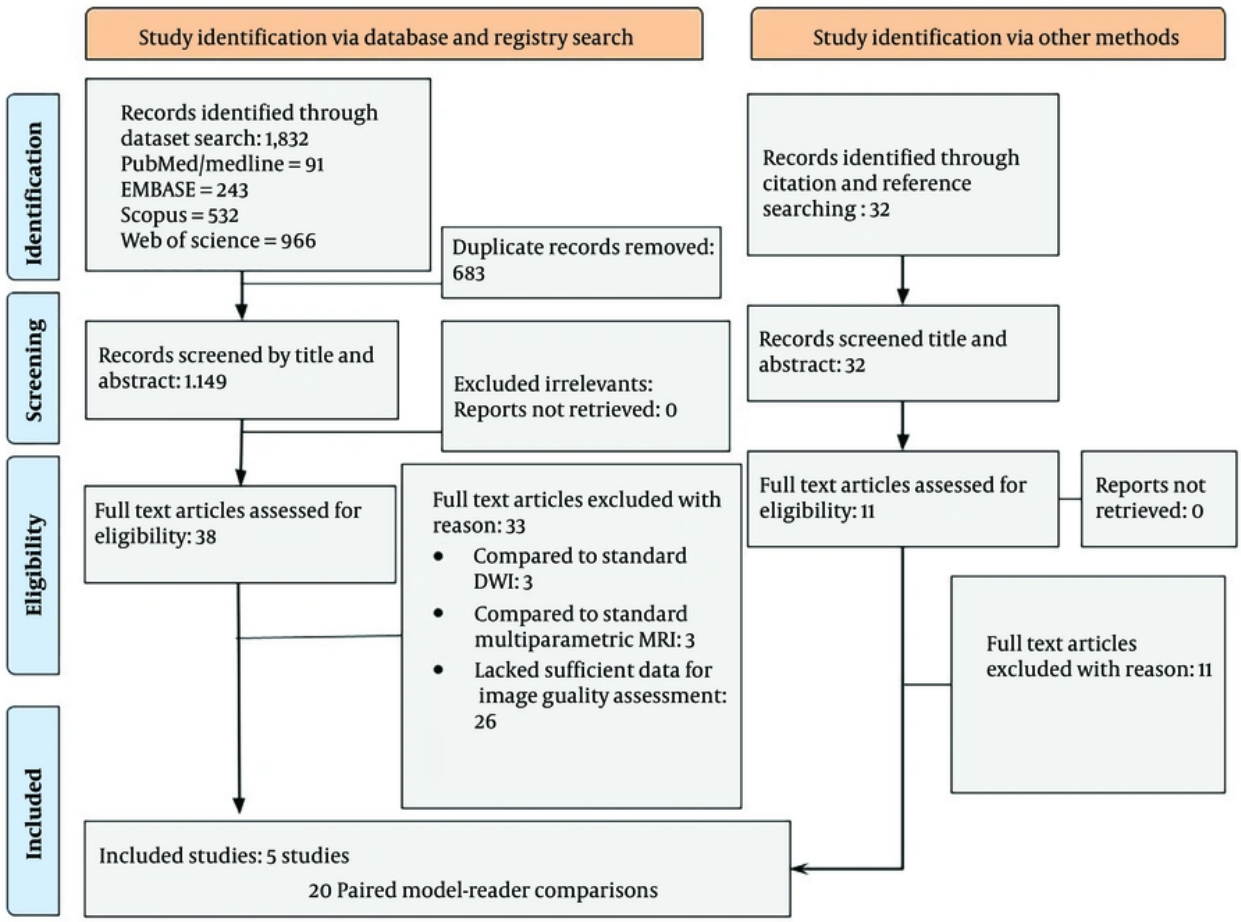

Figure 1. PRISMA 2020 flow diagram.

4.4. Acceleration Meta-Regression

Acceleration Meta-Regression: Including acquisition-time acceleration in our three-level meta-regression revealed a statistically significant trade-off between faster imaging and image quality (Figure 3). At an acceleration of 1.0, the pooled SMCC was estimated at 2.21 SD (95% CI -0.17 to 4.58; P = 0.069), indicating a non-significant trend toward improved image quality under near-isochronous conditions. For each fold increase in

Table 2. Technical Settings and Results Reported in the Included Studies

Study, Year	Deep Learning Reconstruction	Reference	Vendor	SNR (model / ref)	Scan Time (s), Model / Ref	Overall Image Quality (DL Model / Ref)	Likert Scale
Liu et al. (22)	CycleGAN super-res. T2-SR	Standard T2W-TSE	Siemens	N/A	75 / 750	R1: 3.11 ± 0.41 / 3.81 ± 0.42	4
Johnson et al. (23)	Variational-Network DL	GRAPPA (parallel)	Siemens	N/A	42 / 232.2	R1: 3.90 ± 0.64 / 4.00 ± 0.56; R2: 3.80 ± 0.89 / 4.35 ± 0.74; R3: 4.55 ± 0.60 / 4.60 ± 0.50; R4: 3.60 ± 1.00 / 3.65 ± 0.99	5
Harder et al. (21)	C-SENSE AI 1.7	C-SENSE 1.7	Philips	7.68 ± 1.50 / 4.30 ± 0.47	285 / 285	Mean of 4 readers: 5.06 ± 0.79 / 4.60 ± 1.17	6
Harder et al. (21)	C-SENSE AI 3.4	C-SENSE 1.7	Philips	6.61 ± 1.78 / 4.30 ± 0.47	156 / 285	Mean of 4 readers: 5.34 ± 0.69 / 3.05 ± 0.76	6
Harder et al. (21)	C-SENSE AI 4.8	C-SENSE 1.7	Philips	5.79 ± 0.72 / 4.30 ± 0.47	119 / 285	Mean of 4 readers: 4.28 ± 0.51 / 3.05 ± 0.76	6
Tong et al. (18)	Variational-Network DL	Conventional Cartesian T2-TSE	Siemens	N/A	68 / 226	R1: 3.89 ± 0.39 / 3.72 ± 0.53; R2: 3.31 ± 0.74 / 3.33 ± 0.82; R3: 3.51 ± 0.62 / 3.67 ± 0.63	4
Lee et al. (25)	AIR Recon DL - Standard, Low denoising	T2WI SoC (DLR off)	GE Healthcare	Median 12.38 (IQR 3.11) / Median 10.07 (IQR 2.12)	362 / 362	3.42 ± 0.98 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - Standard, Medium denoising	T2WI SoC (DLR off)	GE Healthcare	Median 14.24 (IQR 4.17) / Median 10.07 (IQR 2.12)	362 / 362	3.35 ± 1.14 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - Standard, High denoising	T2WI SoC (DLR off)	GE Healthcare	Median 17.28 (IQR 6.49) / Median 10.07 (IQR 2.12)	362 / 362	3.12 ± 1.07 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - Fast, Low denoising	T2WI SoC (DLR off)	GE Healthcare	Median 11.65 (IQR 3.47) / Median 10.07 (IQR 2.12)	244 / 362	3.58 ± 1.01 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - Fast, Medium denoising	T2WI SoC (DLR off)	GE Healthcare	Median 14.13 (IQR 4.74) / Median 10.07 (IQR 2.12)	244 / 362	3.78 ± 0.89 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - Fast, High denoising	T2WI SoC (DLR off)	GE Healthcare	Median 18.70 (IQR 6.74) / Median 10.07 (IQR 2.12)	244 / 362	3.38 ± 0.95 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - High-resolution, Low denoising	T2WI SoC (DLR off)	GE Healthcare	Median 10.57 (IQR 2.87) / Median 10.07 (IQR 2.12)	244 / 362	3.30 ± 0.88 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - High-resolution, Medium denoising	T2WI SoC (DLR off)	GE Healthcare	Median 10.57 (IQR 3.61) / Median 10.07 (IQR 2.12)	244 / 362	3.40 ± 0.98 / 3.23 ± 0.92	5
Lee et al. (25)	AIR Recon DL - High-resolution, High denoising	T2WI SoC (DLR off)	GE Healthcare	Median 13.17 (IQR 4.43) / Median 10.07 (IQR 2.12)	244 / 362	3.12 ± 0.76 / 3.23 ± 0.92	5

Abbreviations: DL, deep learning; SNR, signal-to-noise ratio; R, reader; N/A, not available; CI, confidence interval; IQR, interquartile range; SoC, standard of care; GRAPPA, generalized autocalibrating partially parallel acquisitions; DLR, deep learning reconstruction; CycleGAN, GAN-based super-resolution models; C-SENSE AI, compressed sensing combined with AI; SR, super resolution; T2WI, T2 weighted imaging; TSE: turbo spin echo; Ref, reference; Recon, reconstruction.

acceleration, image quality declined by 0.48 SD (95% CI -0.91 to -0.05; P = 0.030). Substantial residual heterogeneity persisted (QE18 = 399.5, P < 0.0001). Given the limited number of included effects, these findings should be interpreted as exploratory.

4.5. Risk-of-Bias Analysis

Risk-of-Bias Analysis Visual inspection of the funnel plot, together with PET line (Figure 4), suggests the presence of small-study effects for Image-Quality. The PET slope (β_1 = 8.86, P < 0.0001) shows a positive association between study-level sampling error and reported effect size, indicating that smaller studies tend to report larger estimates of deep learning image quality.

5. Discussion

Deep learning reconstruction evaluations have largely focused on image quality and methodological heterogeneity and variability in quality metrics restricted quantitative comparisons. However, our study indicates that, in prostate imaging, deep learning reconstruction neither significantly improves nor degrades the overall image quality of T2-weighted prostate MRI compared with conventional reconstruction, while enabling reductions in scan time. Meta-regression analysis demonstrated that faster deep

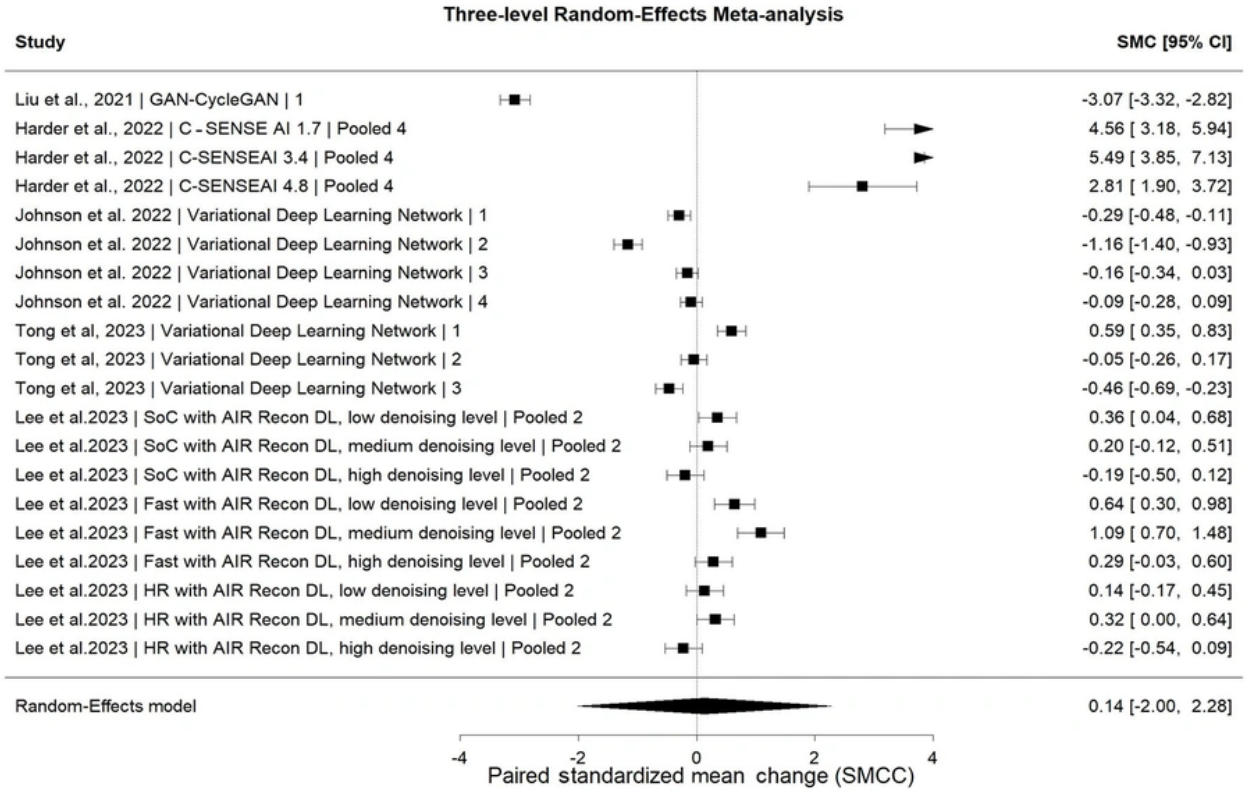

Figure 2. Forest plot of the paired standardized mean change (SMCC) in overall image quality comparing deep-learning versus standard T2-weighted reconstructions across different models and readers. Squares show individual study effects (size proportional to study weight); horizontal bars represent 95% confidence intervals (CIs). The diamond indicates the overall pooled SMCC, and the vertical line at 0 indicates no difference in image quality-values to the right favor deep learning reconstruction. SoC: Standard of Care; HR: High Resolution; DL, deep learning.

learning models compared to conventional models are associated with significant decline in image quality.

These results are clinically important given the increasing need to shorten MRI examination times (18). Nevertheless, a critical question in deep learning models implementation is whether faster models come at the cost of image quality and finally diagnostic accuracy. Our meta-regression between acceleration and image quality differences showed that faster deep-learning models relative to conventional reconstruction were significantly associated with reduced image quality, corresponding to a decline of 0.48 SD for each one-fold increase in speed. Importantly, the observed heterogeneity necessitates cautious interpretation of these results. A previous study showed that radiologists can comfortably use deep learning acceleration in the range of 2-3 for routine imaging without decreasing diagnostic confidence (19). Ursprung et al. study suggests deep learning reconstruction can shorten prostate MRI acquisitions while often preserving image quality, but the evidence base is highly heterogeneous across sequences, vendors, study designs (20). Radiologists thus face a dilemma in choosing the acceleration, balancing patient burden and comfort against the need for anatomic details. Clinical teams need to create models with the acceptable acceleration level based on their aim. As an example, using higher acceleration where some image blurring is acceptable (screening or routine follow-up) and using lower acceleration when detail must be visualized.

A significant concern in deep learning reconstruction is the occurrence of hallucinations, where realistic-looking but incorrect image details are added or removed. This issue arises when there are multiple plausible solutions to the inverse image reconstruction problem. A model, influenced by the

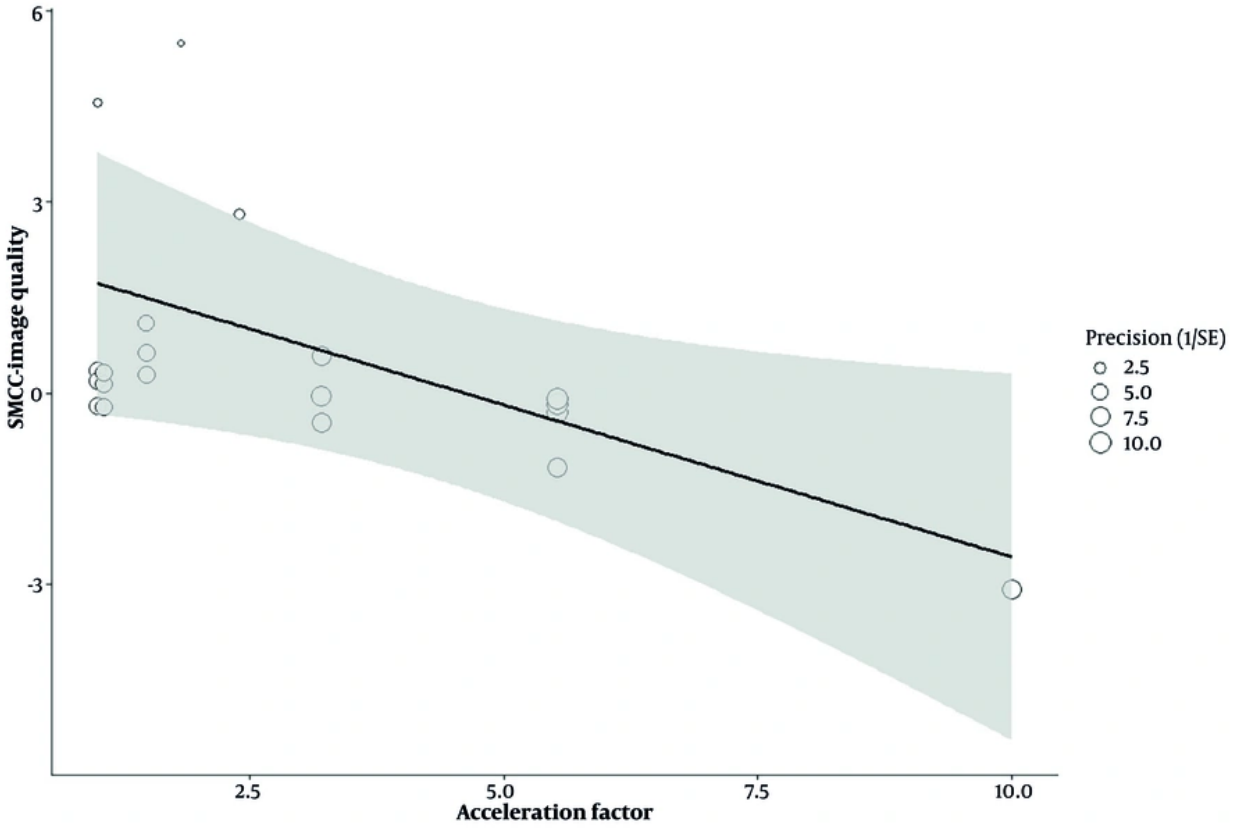

Figure 3. Bubble plot illustrating the meta-regression of scan-time acceleration versus paired standardized mean change (SMCC) in overall image quality.

patterns learned from training data, may favor solutions that introduce fabricated features (20). Deep learning models can also cause over-smoothing, particularly in the transition zone, which may mimic the charcoal sign, leading to false positives or lost subtle findings near the capsule, resulting in false negatives (20). The risk of hallucinations and artifacts increases with stronger acceleration and decreased sampling rates. Artifacts can also vary based on algorithm settings, vendors, scanners, and patient anatomy (20).

Among the models assessed, C-SENSE AI by Harder et al, delivered the best image quality over their respective conventional models (21). In contrast, GAN with CycleGAN for super-resolution reconstruction degraded image quality compared to its conventional models (22). These differences can be traced to balances in under-sampling, noise suppression, detail preservation, and acceleration, pointing the need for careful selection and validation of deep learning models in clinical setting. C-SENSE AI paired modest k-space under-sampling with a multiscale CNN embedded directly in the raw-data (21). They kept the acceleration within a range that preserved SNR and concentrated on denoising and edge sharpening rather than hallucinating missing anatomy (21). By contrast, architectures that pushed acceleration or super-resolution too far fared poorly. A CycleGAN super-resolution model up-sampled thick 2-D T2-TSE slices to 0.75 mm isotropic but introduced striping, hallucinated textures, and uneven high-frequency noise, leading readers to rate it below the conventional model (22). In evaluated underperformed models, resolution or speed was gained at the expense of raw SNR and new artifacts the network could not fully correct. Vendor-specific training limits how well they generalize, and radiologists still encounter over-smoothed or noisy images even when metrics look good, so each center should validate any deep learning tool on its own scanners, especially in thin-slice or under-sampled protocols where weaknesses emerge first.

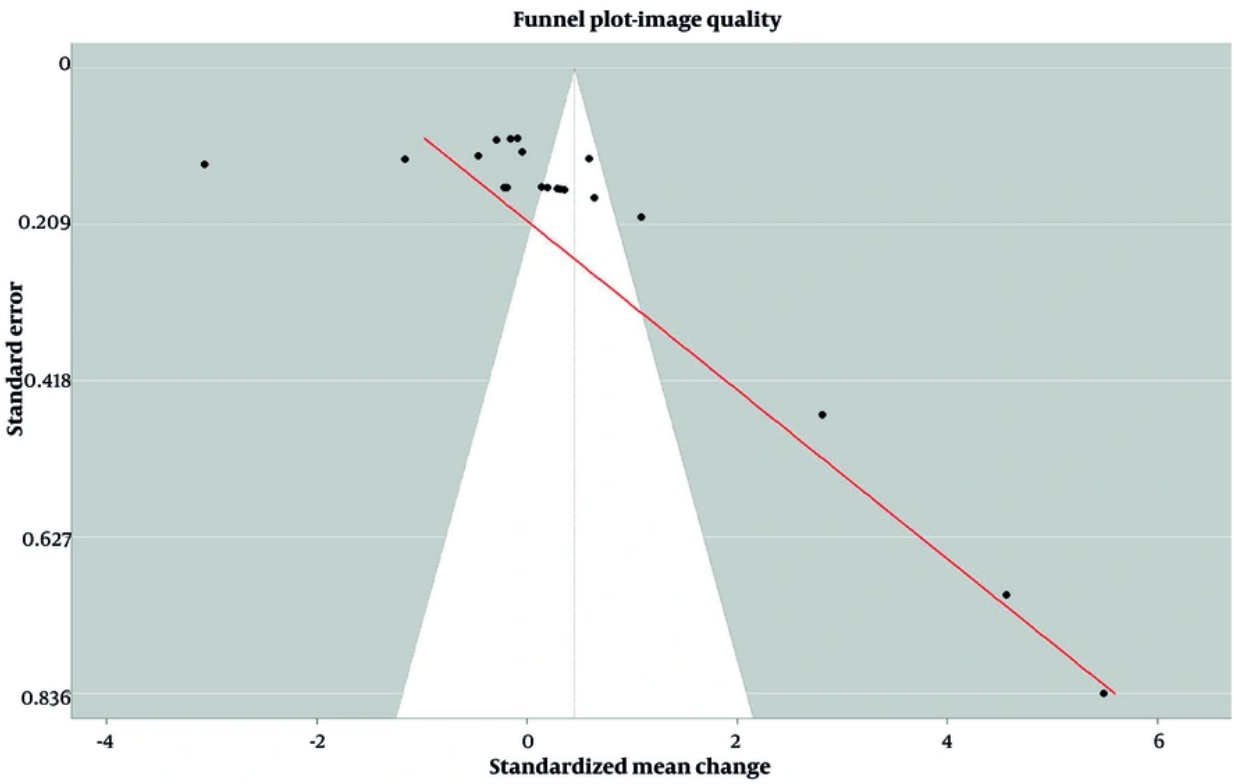

Figure 4. Funnel plot with precision-effect test (PET) line assessing small-study effects.

Our findings provide several implications for radiologists and clinical practice. Firstly, a radiologist can expect that a well-trained deep learning reconstruction algorithm will preserve diagnostic information present in a conventional T2-weighted image. Secondly, the ability to shorten MRI exams without loss of information directly benefits patient care. Prostate MRI, which traditionally could take 30-45 minutes for a full multiparametric study, may be shorten into a sub-10-minute with deep learning acceleration (23-24). Shorter scan times translate to less time that patients must remain in the scanner, alleviating discomfort (especially important for elderly patients or those with pain) and reducing anxiety for claustrophobic individuals. They also diminish the likelihood of motion artifacts, since there is a smaller window during which the patient might move. By increasing throughput, faster exams allow more patients to be scanned per day, which can help address long MRI waiting lists and improve access to prostate MRI as a screening or diagnostic tool. However, our findings also provide a note of caution. Clinicians need to be aware that deep learning reconstructions may sometimes introduce changes to image appearance (smoothing, minor artifacts) that differ from conventional images. Training and experience are needed to recognize normal vs. abnormal after deep learning processing. There can be a learning curve which radiologists must get comfortable with the DL-enhanced images and ensure they detect any failure case.

This study is not without limitations. First, the number of studies available for inclusion was relatively small (five studies with 20 comparisons), showing the early stage of this technology's evaluation. The high heterogeneity we observed indicates that results varied widely between settings. We addressed some of this by meta-regressions on acceleration, but other sources of heterogeneity likely remain, such as how conventional reconstruction was defined. Second, we observed asymmetry in funnel plots, with smaller studies tending to report larger deep learning benefits. Third, our meta-analysis focused on image quality rather than direct

patient outcomes. Future research should evaluate clinical end-points such as diagnostic accuracy, sensitivity, and specificity. Fourth, the rapid evolution of deep learning models means our meta-analysis is a moving target. It will be important to periodically update analyses to include newer studies and possibly head-to-head comparisons of different deep learning approaches. Multi-center trials with a bulk of patients, randomizing or crossing over between deep learning and conventional reconstructions, could provide stronger evidence of diagnostic performance and capture fewer failure modes. Likert-type scales with ≥4 ordered categories were treated as semi-continuous measures, while acknowledging that the assumption of equal spacing between adjacent categories may not strictly hold. Assessment of small-study effects using the Precision-Effect Test (PET) was limited by the small number of included studies. As PET has low statistical power when fewer than ten studies are available, these results should be interpreted with caution and viewed as exploratory rather than confirmatory.

To put these findings into clinical context, we assessed the certainty of evidence using a GRADE framework. The overall certainty for image-quality outcomes was rated as low. This rating reflects the substantial inter-study heterogeneity, the use of surrogate image-quality metrics instead of direct clinical outcomes (indirectness), and the limited precision resulting from the small number of included studies. Therefore, while deep learning reconstruction shows promise for scan-time acceleration, our results should be interpreted as exploratory rather than definitive.

Footnotes

AI Use Disclosure: During the preparation of this work, the authors used ChatGPT (OpenAI, San Francisco, USA) for English language refinement and structural editing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final content of the manuscript.

Authors' Contribution: Study concept and design: S.D. Acquisition of data: H.B. and M.N.S. Analysis and interpretation of data: N.A. Drafting of the manuscript: N.A. and H.B. Critical revision of the manuscript for important intellectual content: S.D. Statistical analysis: N.A. Administrative, technical, and material support: H.B. Study supervision: S.D.

Conflict of Interests Statement: 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.

Data Availability: The data presented in this study are uploaded during submission as a supplementary file and are openly available for readers upon request.

Funding/Support: The authors declare that no funding or financial support was obtained for the design, execution, analysis, or publication of this study.

References

1. Rawla P, et al. Epidemiology of prostate cancer. *World J Oncol*. 2019;**10**(2):63. [PubMed ID: 31068988]. [PubMed Central ID: PMC6497009]. https://doi.org/10.14740/wjon1191.
2. Wu L-M, et al. The clinical value of diffusion-weighted imaging in combination with T2-weighted imaging in diagnosing prostate carcinoma: a systematic review and meta-analysis. *Am J Roentgenol*. 2012;**199**(1):103-110. [PubMed ID: 22733900]. https://doi.org/10.2214/AJR.11.7634.
3. Gassenmaier S, et al. Thin-slice prostate MRI enabled by deep learning image reconstruction. *Cancers*. 2023;**15**(3):578. [PubMed ID: 36765539]. [PubMed Central ID: PMC9913660]. https://doi.org/10.3390/cancers15030578.
4. Hoffman RM, et al. Screening for prostate cancer. *N Engl J Med*. 2011;**365**(21):2013-2019. [PubMed ID: 22029754]. https://doi.org/10.1056/NEJMcp1103642.
5. Litwin MS, et al. The diagnosis and treatment of prostate cancer: a review. *JAMA*. 2017;**317**(24):2532-2542. [PubMed ID: 28655021]. https://doi.org/10.1001/jama.2017.7248.
6. Li H, et al. Machine Learning in Prostate MRI for Prostate Cancer: Current Status and Future Opportunities. *Diagnostics*. 2022;**12**(2):289. [PubMed ID: 35204380]. [PubMed Central ID: PMC8870978]. https://doi.org/10.3390/diagnostics12020289.
7. Ueda T, et al. Deep Learning Reconstruction of Diffusion-weighted MRI Improves Image Quality for Prostatic Imaging. *Radiology*. 2022;**303**(2):373-381. [PubMed ID: 35103536]. https://doi.org/10.1148/radiol.204097.
8. Cho J, et al. Biparametric versus multiparametric magnetic resonance imaging of the prostate: detection of clinically significant cancer in a perfect match group. *Prostate Int*. 2020;**8**(4):146-151. [PubMed ID: 33425791]. [PubMed Central ID: PMC7767942]. https://doi.org/10.1016/j.prnil.2019.12.004.
9. Bischoff LM, et al. Deep Learning Super-Resolution Reconstruction for Fast and Motion-Robust T2-weighted Prostate MRI. *Radiology*. 2023;**308**(3):e230427. [PubMed ID: 37750774]. https://doi.org/10.1148/radiol.230427.
10. Cuocolo R, et al. Machine learning applications in prostate cancer magnetic resonance imaging. *Eur Radiol Exp*. 2019;**3**(1):35. [PubMed ID: 31392526]. [PubMed Central ID: PMC6686027]. https://doi.org/10.1186/s41747-019-0109-2.
11. Gloe JN, et al. Deep learning for quality assessment of axial T2-weighted prostate MRI: a tool to reduce unnecessary rescanning. *Eur Radiol Exp*. 2025;**9**(1):44. [PubMed ID: 40299162]. [PubMed Central ID: PMC12040773]. https://doi.org/10.1186/s41747-025-00584-z.
12. Wang G, et al. Deep learning for tomographic image reconstruction. *Nat Mach Intell*. 2020;**2**(12):737-748. https://doi.org/10.1038/s42256-020-00273-z.
13. Antun V, et al. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proc Natl Acad Sci U S A*.

2020;**117**(48):30088-30095. [PubMed ID: 32393633]. [PubMed Central ID: PMC7720232]. https://doi.org/10.1073/pnas.1907377117.

14. van Lohuizen Q, et al. Assessing deep learning reconstruction for faster prostate MRI: visual vs. diagnostic performance metrics. *Eur Radiol*. 2024;**34**(11):7364-7372. [PubMed ID: 38724765]. [PubMed Central ID: PMC11519109]. https://doi.org/10.1007/s00330-024-10771-y.
15. Moher D, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;**8**(5):336-341. [PubMed ID: 20171303]. https://doi.org/10.1016/j.ijsu.2010.02.007.
16. Page MJ, et al. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. *Syst Rev*. 2017;**6**(1):263. [PubMed ID: 29258593]. [PubMed Central ID: PMC5738221]. https://doi.org/10.1186/s13643-017-0663-8.
17. Whiting P, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;**3**:25. [PubMed ID: 14606960]. [PubMed Central ID: PMC305345]. https://doi.org/10.1186/1471-2288-3-25.
18. Tong A, et al. Comparison of a Deep Learning-Accelerated vs. Conventional T2-Weighted Sequence in Biparametric MRI of the Prostate. *J Magn Reson Imaging*. 2023;**58**(4):1055-1064. [PubMed ID: 36651358]. [PubMed Central ID: PMC10352465]. https://doi.org/10.1002/jmri.28602.
19. Foti G, et al. Deep learning and AI in reducing magnetic resonance imaging scanning time: advantages and pitfalls in clinical practice. *Pol J Radiol*. 2024;**89**:e443-e451. [PubMed ID: 39444654]. [PubMed Central ID: PMC11497590]. https://doi.org/10.5114/pjr/192822.
20. Ursprung S, et al. Prostate MRI Using Deep Learning Reconstruction in Response to Cancer Screening Demands-A Systematic Review and Meta-Analysis. *J Pers Med*. 2025;**15**(7):284. [PubMed ID: 40710401]. [PubMed Central ID: PMC12298121]. https://doi.org/10.3390/jpm15070284.
21. Harder FN, et al. Prospectively Accelerated T2-Weighted Imaging of the Prostate by Combining Compressed SENSE and Deep Learning in Patients with Histologically Proven Prostate Cancer. *Cancers*. 2022;**14**(23):5741. [PubMed ID: 36497223]. [PubMed Central ID: PMC9738899]. https://doi.org/10.3390/cancers14235741.
22. Liu YC, et al. 3D Isotropic Super-resolution Prostate MRI Using Generative Adversarial Networks and Unpaired Multiplane Slices. *J Digit Imaging*. 2021;**34**(5):1199-1208. [PubMed ID: 34519954]. [PubMed Central ID: PMC8555005]. https://doi.org/10.1007/s10278-021-00510-w.
23. Johnson PM, et al. Deep Learning Reconstruction Enables Highly Accelerated Biparametric MR Imaging of the Prostate. *J Magn Reson Imaging*. 2022;**56**(1):184-195. [PubMed ID: 34877735]. [PubMed Central ID: PMC9170839]. https://doi.org/10.1002/jmri.28024.
24. Girometti R, et al. Evolution of prostate MRI: from multiparametric standard to less-is-better and different-is better strategies. *Eur Radiol Exp*. 2019;**3**(1):5. [PubMed ID: 30693407]. [PubMed Central ID: PMC6890868]. https://doi.org/10.1186/s41747-019-0088-3.
25. Lee KL, et al. Assessment of deep learning-based reconstruction on T2-weighted and diffusion-weighted prostate MRI image quality. *Eur J Radiol*. 2023;**166**:111017. [PubMed ID: 37541181]. https://doi.org/10.1016/j.ejrad.2023.111017.