

# The application of metal artifact reduction methods on computed tomography scans for radiotherapy applications: A literature review

Sathyathas Puvanasuntharajah<sup>1,2</sup> | Davide Fontanarosa<sup>1,2</sup> | Marie-Luise Wille<sup>2,3,4</sup> | Saskia M. Camps<sup>5</sup>

<sup>1</sup>School of Clinical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

<sup>2</sup>Centre for Biomedical Technologies, Queensland University of Technology, Brisbane, QLD, Australia

<sup>3</sup>School of Mechanical, Medical & Process Engineering, Faculty of Engineering, Queensland University of Technology, Brisbane, QLD, Australia

<sup>4</sup>ARC ITTC for Multiscale 3D Imaging, Modelling, and Manufacturing, Queensland University of Technology, Brisbane, QLD, Australia

<sup>5</sup>EBAMed SA, Geneva, Switzerland

Author to whom correspondence should be addressed. Saskia Camps  
Email: saskia.camps@eba-med.com;  
Telephone: +41 78 687 86 20.

## Abstract

Metal artifact reduction (MAR) methods are used to reduce artifacts from metals or metal components in computed tomography (CT). In radiotherapy (RT), CT is the most used imaging modality for planning, whose quality is often affected by metal artifacts. The aim of this study is to systematically review the impact of MAR methods on CT Hounsfield Unit values, contouring of regions of interest, and dose calculation for RT applications. This systematic review is performed in accordance with the PRISMA guidelines; the PubMed and Web of Science databases were searched using the main keywords “metal artifact reduction”, “computed tomography” and “radiotherapy”. A total of 382 publications were identified, of which 40 (including one review article) met the inclusion criteria and were included in this review. The selected publications (except for the review article) were grouped into two main categories: commercial MAR methods and research-based MAR methods. Conclusion: The application of MAR methods on CT scans can improve treatment planning quality in RT. However, none of the investigated or proposed MAR methods was completely satisfactory for RT applications because of limitations such as the introduction of other errors (e.g., other artifacts) or image quality degradation (e.g., blurring), and further research is still necessary to overcome these challenges.

## KEY WORDS

computed tomography, metal artifact reduction, metal artifacts, radiotherapy

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Applied Clinical Medical Physics* published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine

**Abbreviations:** AAA, anisotropic analytical algorithm; ALIR, augmented likelihood image reconstruction; AXB, Acuros External Beam; CNNs, convolutional neural network; CT, computed tomography; CT<sub>art</sub>, CT scan with metal artifacts; CT<sub>cor</sub>, CT scan after metal artifact reduction; CT<sub>ref</sub>, CT scan without metal artifacts; DAH, dose area histogram; DL-MAR, deep learning-based metal artifact reduction; DSC, Dice similarity coefficient; GANs, generative adversarial networks; GTV, gross tumor volume; H&N, head and neck; HD, Hausdorff distance; HU, Hounsfield units; iMAR, iterative metal artifact reduction; IMRT, intensity-modulated radiation therapy; kVCT, kilovoltage computed tomography; kVp, kilovoltage peak; LI, linear interpolation; MAR, metal artifact reduction; MBIR, model-based iterative reconstruction; MDT, metal deletion technique; MV, megavoltage; MVCBCT, megavoltage con-beam computed tomography; MVCT, megavoltage computed tomography; N-MAR, normalized metal artifact reduction; OARs, organs at risk; O-MAR, orthopedic metal artifact reduction; PSNR, peak signal-to-noise ratio; PTV, planning target volume; RL-ARCNN, residual learning-artifact reduction convolutional neural network; ROI, region of interest; RT, radiotherapy; SBRT, stereotactic body radiation therapy; SEMAR, single energy metal artifact reduction; SSIM, structural similarity; STD, standard deviation; STR, simple threshold replacement; TPS, treatment planning system; US, ultrasound; VMAT, volumetric arc therapy; WET, water equivalent thickness; %GP, gamma passing rate

## 1 | INTRODUCTION

Radiotherapy (RT) is one of the primary curative treatment options for different types of cancer, for example, cancers of head and neck, prostate, cervix, breast, as well as sarcomas. RT aims to deliver therapeutic ionizing radiation dose to a treatment target, while sparing healthy organs at risk (OARs) as much as possible. The therapeutic radiation dose is delivered to the target using beams produced by a clinical linear accelerator (LINAC) in external beam radiation therapy (EBRT), while in brachytherapy radioactive sources invasively placed near or inside the target are used. Typically, RT workflows include a simulation stage and a treatment delivery stage. During the treatment planning process at the simulation stage, computed tomography (CT) scans serve as a primary source of anatomical information to identify and delineate the target and OARs. In addition, they are used to calculate the electron densities which are derived from the Hounsfield unit (HU) values of those CT scans. This electron density information in combination with the delineations of the anatomical structures is used to calculate the therapeutic radiation dose.

CT scans with insufficient quality may greatly affect the treatment planning process, potentially resulting in the target receiving insufficient dosage and/or extra toxicity to the OARs. Metal implants or metal components inside the body of the patient can induce errors during the CT reconstruction, which appear as artifacts on the resulting CT scans. These metal artifacts are typically bright and/or dark streaks (see Fig. 1) and are produced by beam hardening, photon starvation, edge gradient effect, scatter, or their combination.<sup>1,2</sup> The degree of metal artifacts mainly depends on the atomic number, density, size, and shape of this metal component as well as its orientation with respect to the CT scan plane.<sup>3,4</sup> Among others, dental implants or dental fillings in the head and neck (H&N) area, bilateral or unilateral metal prostheses in the hip region, and metal screws in the spine produce a large amount of metal artifacts and, thus, significantly deteriorate the quality of CT scans.<sup>5-7</sup>

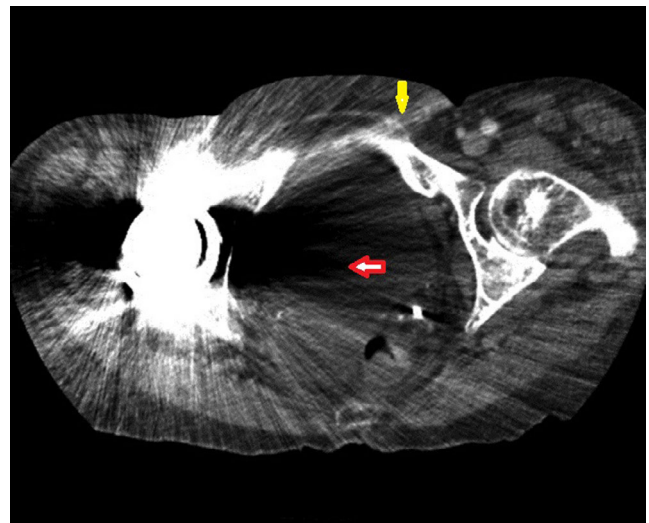
Dark streaks near the metal components result from highly attenuated polychromatic x-ray beams, which become for this reason harder.<sup>8</sup> Because of this, insufficient photons reach the CT detectors (photon starvation), resulting in large statistical errors in data acquisition, which induce fine bright and dark streaks along the direction of highest attenuation<sup>7,8</sup>. As a consequence, the appearance of these streak-shaped metal artifacts adversely affects the accuracy of organ contouring and the electron density calculation. This can eventually result in errors in planned radiation dose distributions and particle range measurements in photon and particle beams, respectively.<sup>9,10</sup>

In the literature, several papers have been published on algorithms which perform metal artifact reduction (MAR) on CT scans. The working principle of traditional MAR algorithms may be categorized into three overall approaches: image inpainting techniques,<sup>11</sup> sinogram inpainting techniques,<sup>12</sup> and model-based iterative reconstruction (MBIR) techniques<sup>13</sup> or their combination. The image inpainting techniques are applied to already reconstructed CT scans and they replace artifact corrupted CT pixels with good-estimated values. The sinogram inpainting techniques follow a similar principle,

but are used on projection data (sinograms) instead of on reconstructed CT slices. Finally, MBIR techniques are advanced CT reconstruction techniques which use probabilistic forward and backward models to reduce error propagations during CT reconstruction.<sup>14</sup> Recently, thanks to the increasing availability of computational resources, very promising results in the field of medical imaging have been produced using machine learning (and in particular its subset deep learning),<sup>15-18</sup> including metal artifact reduction in CT scans.<sup>19,20</sup> For example, the performance of convolutional neural networks (CNNs) has been assessed in combination with sinogram inpainting for artifact correction.<sup>19,21</sup> The deep learning techniques are powerful in learning and capturing the detailed features and patterns of the metal artifacts.

In general, the application of a MAR method on a CT scan with artifacts ( $CT_{art}$ ) results in the creation of a corrected CT scan ( $CT_{cor}$ ) on which the impact of the artifacts is reduced, either in terms of image quality or dosimetric outcome on the treatment. To measure the effectiveness of the methods, several different metrics have been introduced in the literature to compare  $CT_{art}$  and  $CT_{cor}$ . Image quality metrics proposed include visual inspection, quantification of HU values, artifact index,<sup>22</sup> contrast-to-noise ratio (CNR), signal-to-noise ratio (SNR), peak SNR (PSNR), structural similarity (SSIM), Hausdorff distance (HD),<sup>23</sup> and the Dice similarity coefficient (DSC).<sup>24-26</sup> To evaluate the dosimetric impact, instead, the calculated dose distributions on  $CT_{art}$  and  $CT_{cor}$  for the target and OARs provided by a treatment planning system (TPS) can be compared. Various dose metrics can be used to express the dosimetric impact, including gamma ( $\gamma$ ) index,<sup>27,28</sup> dose-area histogram (DAH),<sup>29</sup> quantifications of  $D_{90\%}$ ,  $D_{100\%}$ ,  $V_{100\%}$ , and  $V_{150\%}$ ,<sup>30</sup> and therapeutic range calculation (in particular, water equivalent thickness (WET)<sup>31</sup> in particle therapy).

A topical review article has been previously published by Giant-soudi et al. on the evaluation of the dosimetric effects of metal



**FIG. 1.** CT scan with artifacts induced by a unilateral hip implant with the appearance of bright streaks (indicated by yellow arrow) and dark streaks (indicated by red arrow).

artifacts on treatment planning and the potential dosimetric improvements resulting from the application of various MAR methods.<sup>10</sup> The article addressed the impact of sinogram inpainting and MBIR on the dose distributions, mainly focusing on research-based MAR methods while, among the commercially available MAR methods, only the Orthopaedics Metal Artifact Reduction (O-MAR (Philips Health System, Cleveland, USA)) algorithm was reported. Instead, this systematic review article aims to include all MAR methods which have been investigated or proposed for RT applications in the last 5 yrs at the time of publication (2015–2020). These methods include commercial MAR methods and research-based MAR methods based on either traditional algorithms or deep learning. In addition, our review extensively reports not only the works on dosimetric impact of the methods but also on the ones evaluating the effects on organ contouring, and image quality and HU restoration for RT applications.

## 2 | METHODS

### 2.A | Literature search

The systematic review search was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.<sup>32,33</sup> A comprehensive electronic search from the databases of PubMed® (U.S. National Library of Medicine, USA) and Web of Science (Clarivate Analytics, USA) was performed in October 2020. Combinations and synonyms of the main keywords “metal artifact reduction,” “computed tomography,” and “radiotherapy” were used. The search was limited to the English language and to the last 5 yrs. Initially, title and abstract of the identified articles were read to screen their suitability for the selection. Then, the full texts were read from the selected articles to check their eligibility for inclusion. Finally, a manual search was performed using the list of references of the included articles to find any additional data missed by the initial database searches (see Fig. 2).

### 2.B | Inclusion and exclusion criteria

An article was considered if it investigated the use of one or multiple MAR methods on CT scans for RT applications, with the exclusion of dual-energy CT (DECT), dental cone-beam CT (CBCT), C-arm CT, spectral CT, micro-CT, or photoacoustic CT. Also, editorial commentaries and book chapters were excluded.

## 3 | RESULTS

A total of 40 full-text publications were selected for this systematic review, including one review article, as mentioned in the Introduction section. The selected publications (except the review article) have further been categorized into application of commercial methods ( $n = 25$ ), and application of research-based MAR methods ( $n = 14$ ). The category of commercial methods includes the articles on commercial MAR algorithms ( $n = 21$ ) and TPS-based density

correction ( $n = 4$ ). Research-based MAR methods include the articles on traditional MAR algorithms ( $n = 11$ ) and deep learning-based MAR algorithms ( $n = 3$ ).

### 3.A | Commercial MAR methods

Commercial MAR methods are available on CT scanners or on the TPS to reduce metal artifacts on  $CT_{art}$  for RT applications. The algorithms implemented directly on CT scanners use a sinogram inpainting technique with iterative reconstruction and include: O-MAR (Philips Health System, Cleveland, USA),<sup>34</sup> iterative metal artifact reduction (iMAR [Siemens Healthcare, Forchheim, Germany]),<sup>35</sup> smart metal artifact reduction (Smart MAR [General Electric Healthcare, Chicago, IL, USA]),<sup>36</sup> and single-energy metal artifact reduction (SEMAR [Canon/Toshiba Medical Systems, Otawara, Japan]).<sup>37</sup> These algorithms work on projection data (projection-based MAR algorithms) and they typically use an image-based metal segmentation method as a starting point.<sup>7</sup> Their basic concept is to detect and segment the corrupted projection data which corresponds to the metal components. Subsequently, the corrupted data are replaced by estimated corrected values.

Several studies have investigated the applicability of commercial MAR algorithms, especially O-MAR, in RT. These commercial MAR algorithms are not openly accessible; however, O-MAR is chosen here for a general introduction to the working principles of all of them. The O-MAR algorithm is optimized for orthopedic metal implants and uses an iterative projection modification method to reduce the artifacts (see Fig. 3). First, a tissue classification process is performed which assigns tissue labels to all the pixels of an original input image (see arrows ‘a’ in Fig. 3). Two separate images are, thus, produced: one including only tissues and one only the metal component. Subsequently, both segmented images (tissue classified and metal only) and the original input image are forward projected to make their respective sinograms (see arrows ‘b’ in Fig. 3). After the sinograms are made, the tissue classified sinogram is subtracted from the original sinogram (see arrows ‘c’ in Fig. 3) resulting in the creation of a difference sinogram. Then, the metal-only sinogram is used as a mask to remove the nonmetal pixels from the difference sinogram (see arrows ‘d’ in Fig. 3) and a mask sinogram is created. A correction image is then produced by filtered back projection of the mask sinogram (see arrow ‘e’ in Fig. 3), and subtraction of the correction image from the original input image is created (see arrows ‘f’ and ‘g’ in Fig. 3). Finally, the corrected image is used as an input in an iterative loop for further correction (see arrow ‘h’ in Fig. 3).

A density correction or a density override method on the TPS can also be used to reduce the metal artifacts on  $CT_{art}$ . When these approaches are used, regions corrupted by metal artifacts and metal regions are identified manually through contouring on  $CT_{art}$ . Then, a manual density override or density correction is performed by replacing the metal artifacts commonly with the density of water or by replacing the physical density of a metal implant by the appropriate value.<sup>38,39</sup>

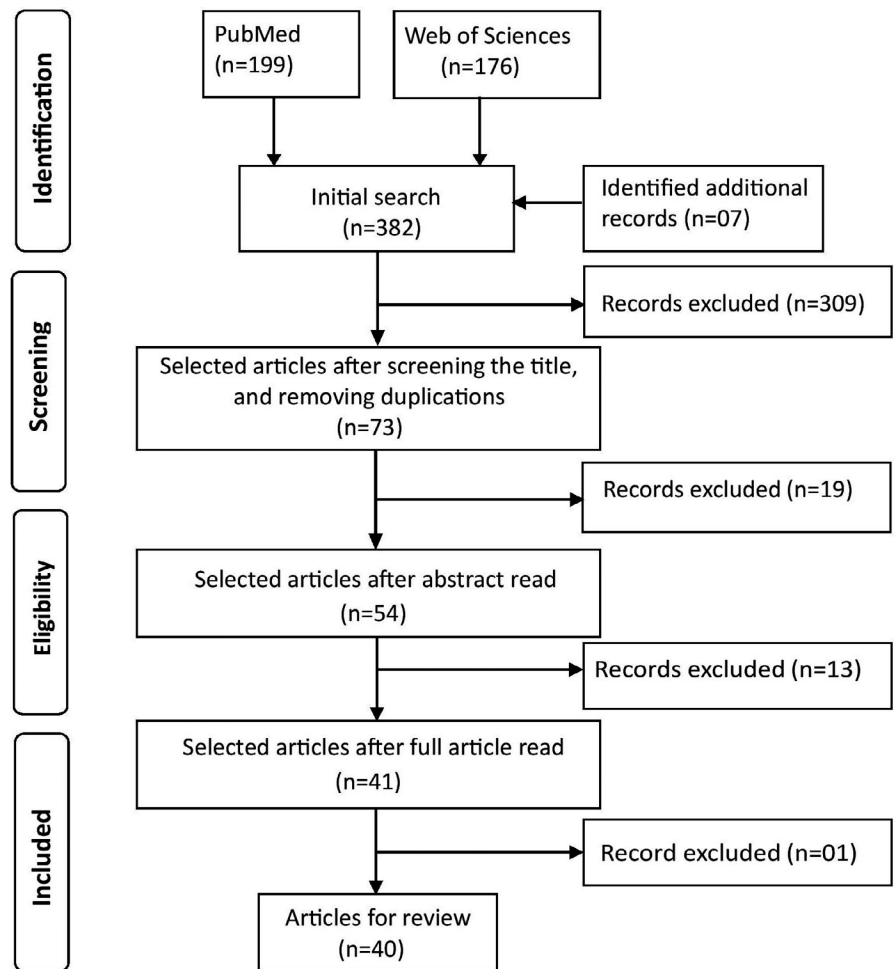


FIG. 2. PRISMA flow chart for the selection of the articles for this review.

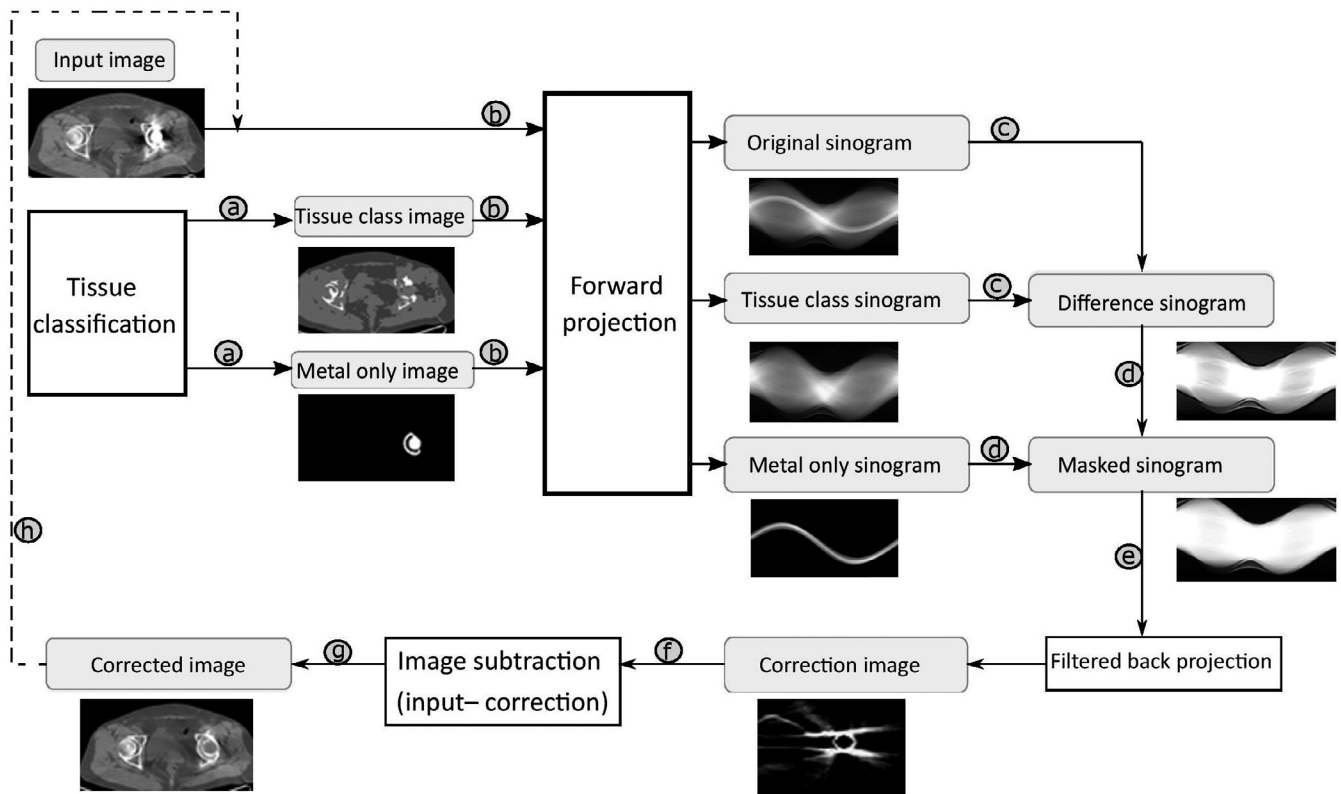
### 3.A.1 | Evaluation of HU values restoration

The capability of the commercial MAR algorithms to improve the HU values for RT applications was evaluated by several studies. Generally, during the HU measurements for an anatomical region on a CT scan, a 10%–20% variations in HU values are expected. For example, at routine CT scans, the HU value of water is set to 0 HU. The HU value of air is approximately –1000 HU and bone is around 1000 HU, while soft tissue HU values range from 20 HU to 30 HU.<sup>40</sup> Table 1 provides an overview of the studies which investigate the HU value improvements.

Kwon et al. studied the impact of O-MAR on both clinical H&N CT scans with dental implants, and on CT scans of a custom-made phantom with aluminum (Al), titanium (Ti), zirconium (Zr), and chromium (Cr) metal implants.<sup>41</sup> The  $CT_{cor}$  after O-MAR application showed HU values closer (not quantified) to the actual values, while the noise on clinical H&N CT scans was reduced. For the phantom, a comparison between  $CT_{cor}$  after O-MAR application and CT scans without metal implants ( $CT_{ref}$ ) did not show significant differences while significant differences ( $P < 0.05$ ) were observed between  $CT_{art}$  and  $CT_{ref}$ .

For the evaluation of iMAR by Bär et al., a Gammex 467 (Gammex, Middleton, WI, USA) phantom was used with several tissue-

equivalent inserts, such as lung, adipose tissue, breast, liver, and bone.<sup>42</sup> The phantom was scanned with and without Al, Ti, and stainless steel (SS) implants. This study revealed that iMAR application improved the HU accuracy on  $CT_{cor}$  compared to  $CT_{art}$ . For example, iMAR approximated the HU values for tissue-equivalent substitutes in the phantom up to  $\pm 44$  HU compared to  $CT_{ref}$ . A similar finding was found by the authors for H&N clinical CT scans with dental fillings and/or implants. Moreover, they stated that iMAR corrected the HU values independently of the metal density. In another study by Axente et al., a standard electron density phantom (CIRS, Model 062MA, Norfolk, VA, USA) was used with multiple inserts, such as plastic water, bone, muscle, adipose tissue, breast, bone, lung, and liver.<sup>35</sup> The phantom with and without SS inserts with varying diameters was CT scanned. It was shown that the  $CT_{cor}$  after iMAR application restored the HU values well and it had absolute differences of less than 25 HU compared to  $CT_{ref}$ . Nevertheless, residual HU errors (not quantified) were observed on the resulting  $CT_{cor}$ . Furthermore, the study mentioned that iMAR restored the HU independently of the metal component's size and configuration. Maerz et al. reported, in their dental cylindrical phantom study, that the use of iMAR reduced the average HU deviation, from 1006 HU on  $CT_{art}$  to 408 HU in the area which included the metal inserts, and from 283 HU on  $CT_{art}$  to 33 HU in tissue areas.<sup>43</sup>



**Fig. 3.** Working principle of the O-MAR algorithm. Starting from tissue classification, O-MAR produces original, tissue class, and metal-only sinograms (a and b). Subtraction (c) of the tissue class sinogram from the original sinogram results in a difference sinogram. Then, the metal-only sinogram is used to mask (d) the difference sinogram and a correction image is produced after filtered back projection (e). Subtraction of the correction image from the input (f and g) produces a corrected image which then undergo (h) in order to apply further corrections.

Guilfoile et al. investigated the ability of Smart MAR to restore the HU values. To this end, a breast expander consisting of a samarium cobalt magnet was placed on an adult male phantom (CIRS, Model ATOM 701-B, Norfolk, VA, USA).<sup>44</sup> The authors compared the mean HU value on  $CT_{ref}$  (not given) with both  $CT_{art}$  and  $CT_{cor}$  after Smart MAR application. The results showed a reduction in the HU value difference from 29 HU to 3 HU in ipsilateral lung, and from 21 HU to 8 HU in bone close to the metal implant, respectively. In addition, small differences in mean HU values were measured in regions that were further away from the implant. In a study with a Catphan® 504 phantom (The Phantom Laboratory, NY, USA) with a SS implant, Huang and Kohli found that Smart MAR substantially reduced the metal artifacts.<sup>45</sup> On average, the standard deviation (STD) was reduced by 9.1 HU on  $CT_{cor}$  after Smart MAR application in comparison with  $CT_{art}$ . Another study used a custom-made water equivalent phantom with lead implants and concluded that Smart MAR application improved the mean HU value in a region of interest (ROI) close to a lead implant from -862 HU on  $CT_{art}$  to -185 HU on  $CT_{cor}$ .<sup>46</sup> This low HU value measurement for water on  $CT_{art}$  (water usually is around 0 HU) resulted from the severe dark streak artifact from the metal implant. The application of iMAR on  $CT_{art}$  reduced the severity of the dark streaks and, therefore, improved the HU value of water.

A study by Murazaki et al. investigated the results of SEMAR application on HU value measurements accuracy.<sup>47</sup> The authors

made use of a standard electron density phantom (CIRS, Model 062A, Norfolk, VA, USA) with and without Ti implants. To simulate the different tissue types, plugs with different densities were inserted into the phantom, such as muscle plugs and soft tissue plugs. They found that, in the muscle plug, the mean HU value on  $CT_{ref}$ ,  $CT_{art}$ , and  $CT_{cor}$  after SEMAR application were 25.9 HU, -281.8 HU, and 26.1 HU, respectively. A similar pattern of HU value measurements was observed in other tissue plugs. In an experiment by Miki et al., an anthropomorphic head phantom was CT scanned with and without the insertion of metal crews.<sup>48</sup> In areas with severe artifacts, HU value measurements on both  $CT_{art}$  and  $CT_{cor}$  after SEMAR application were compared with  $CT_{ref}$ . This resulted in differences of  $-79.5 \pm 97.2$  HU (mean  $\pm$  STD) and  $-1.4 \pm 19.5$  HU (mean  $\pm$  STD), respectively. The HU value of  $CT_{ref}$  was not reported. The  $CT_{cor}$  after SEMAR application brought the HU values closer to the reference values.

In general, the majority of the studies used phantoms to investigate the capability of the commercial MAR algorithm to restore HU values and they concluded that the investigated commercial MAR algorithms improved the accuracy of HU value measurements. Nevertheless, residual HU errors were observed on  $CT_{cor}$ , either underestimating or overestimating the correct HU values. Underestimation resulted from incomplete correction of dark streak artifacts, while overestimation resulted from incomplete correction of bright streaks on  $CT_{cor}$ .

**TABLE 1** Summary of the studies which investigated commercial MAR algorithms for HU value restoration for RT applications.

Author	MAR	Images	Metals	CT scans	ROI; [Reference HU value (Mean)]	Measurements	Results	Findings
Kwon et al. <sup>41</sup>	O-MAR	Clinical (H&N, open mouth), (n = 3)	Dental implants	CT <sub>art</sub> & CT <sub>cor</sub>	Air in outer cavity; [-1000 HU]	Mean HU	-238.7 HU & -441.8 HU	O-MAR increased the accuracy of HU values
Bär et al. <sup>42</sup>	iMAR	Phantom	Al, Ti & SS	CT <sub>ref</sub> vs CT <sub>cor</sub>	Tissue equivalent substitutes; [Not reported]	Absolute HU difference	44 HU	iMAR restored the HU values independent of the metal density
Axente et al. <sup>35</sup>	iMAR	Phantom	SS with varying diameters	CT <sub>ref</sub> vs CT <sub>cor</sub>	Multiple places	Absolute HU difference	25 HU	iMAR restored the HU values independent of the metal size and configuration
Maerz et al. <sup>43</sup>	iMAR	Phantom	Dental implants	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Tissue; [Not reported] Metal inserts; [Not reported]	Mean HU difference	283 HU & 33 HU 1006 & 408 HU	iMAR improved the accuracy of HU value measurements
Guilfoile et al. <sup>44</sup>	Smart MAR	Phantom	Samarium cobalt magnet	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Ipsilateral lung Spinal bone [Not reported]	Mean HU	29HU & 2HU 21HU & 8HU	Application of Smart MAR improved the HU values restoration
Huang & Kohli <sup>45</sup>	Smart MAR	Phantom	SS	CT <sub>cor</sub> vs CT <sub>art</sub>	Multiple places	STD reduction	9.1 HU	Smart MAR reduced the STD of HU values
Inal & Sarpün <sup>46</sup>	Smart MAR	Phantom	Lead	CT <sub>art</sub> & CT <sub>cor</sub>	Area close to the metal edge; [Around 0 HU]	Mean HU	-862 HU & -185 HU	Smart MAR improved the accuracy of HU value measurements
Murazaki et al. <sup>47</sup>	SEMAR	Phantom	Metal inserts	CT <sub>art</sub> & CT <sub>cor</sub>	Muscle substitute; [25.9 HU]	Mean HU	-281.8 HU & 26.1 HU	SEMAR restored the HU values with high level of accuracy
Miki et al. <sup>48</sup>	SEMAR	Phantom	Metal screws	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Severe artifact region; [Not reported]	Mean (± STD) HU	-79.5 (± 97.2) HU & -1.4 (± 19.5) HU	SEMAR improved the HU value measurements accuracy

HU Hounsfield units, H&N Head and neck, Al Aluminum, Ti Titanium, SS Stainless steel, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, vs versus.

### 3.A.2 | Evaluation of organ contouring

The commercial MAR algorithms have been also evaluated on their ability to improve organ contouring in RT (see Table 2). A study by Sillanpaa et al. evaluated the easiness of contouring of the parotid gland on clinical H&N CT scans with dental fillings.<sup>49</sup> The contouring of the parotid was performed on CT<sub>art</sub> and CT<sub>cor</sub> after O-MAR application. During the contouring, 79% of the CT<sub>cor</sub> and 11% of the CT<sub>art</sub> were classified as easy to contour. Furthermore, the authors calculated the dice similarity coefficient (DSC) to assess the

contouring interobserver variability. The perfect match between two contours, on the same anatomy, would result in the highest DSC value of 1, while the lowest DSC value of 0 represents no overlap of two contoured structures. The average DSC for the parotid contouring was  $0.775 \pm 0.045$  (mean  $\pm$  STD) on both CT<sub>cor</sub> (after O-MAR application) and CT<sub>art</sub>. These clinical H&N CT scans had large amounts of small dental fillings on either side of jaw, which might cause incomplete correction of metal artifacts. The authors suggested that this was one of the reasons for classifying 11% of the CT<sub>art</sub> as easy to contour.

**TABLE 2** Summary of the studies which investigated the commercial MAR algorithms for organ contouring for RT.

Author	MAR	Images; No. of sample (n)	Implants	CT scans	ROI	Measurements	Results	Findings
Sillanpaa et al. <sup>49</sup>	O-MAR	Clinical (H&N); (n = 28)	Dental implants	CT <sub>art</sub> & CT <sub>cor</sub> CT <sub>cor</sub> vs CT <sub>art</sub>	Parotid gland	Contouring easiness (%) DSC difference [Mean ± STD]	79% and 11% 0.775 ± 0.045	O-MAR made contouring easy with less interobserver variability
Andersson et al. <sup>50</sup>	O-MAR & iMAR	Clinical (H&N); (n = 30)	Dental implants	CT <sub>art</sub> vs CT <sub>cor</sub>	Whole images	Visual grading score	Refer to fig. 8 in Ref. [50]	O-MAR and iMAR improved the image quality for accurate contouring
Kohan et al. <sup>51</sup>	O-MAR	Clinical (H&N); (n = 11)	Dental implants	CT <sub>art</sub> , CT <sub>cor</sub> & CT <sub>ref</sub>	Masseter muscle, tongue, jaw, upper maxilla, & major pectoralis muscle	ICC All reviewers: Experienced reviewers:	0.884, 0.971 & 0.989 (w/o) 0.903, 0.948 & 0.985 0.934, 0.975 & 0.990 (w/o) 0.904, 0.979 & 0.976	O-MAR application reduced the inter-reader variability of area measurements
Hansen et al. <sup>52</sup>	O-MAR	Clinical (H&N); (n = 11)	Dental implants	CT <sub>art</sub> & CT <sub>cor</sub>	GTV-Tumor GTV-lymph node Parotid	Volume (cm <sup>3</sup> ) [Mean ± STD] DSC of contours [Mean ± STD]	16.1 ± 10.2 & 19.7 ± 13.9 12.5 ± 8.9 & 12.9 ± 9.5 20.0 ± 7.7 & 21.3 ± 7.6 0.60 ± 0.24 & 0.61 ± 0.20 0.75 ± 0.11 & 0.75 ± 0.09 0.70 ± 0.17 & 0.75 ± 0.17	O-MAR improved the volume delineation of structures in a constant way
Hagen et al. <sup>53</sup>	iMAR	Clinical (H&N & hip); (n = 2)	Dental implants and bilateral hip implant	CT <sub>art</sub> & CT <sub>cor</sub> CT <sub>ref</sub> , CT <sub>art</sub> & CT <sub>cor</sub> (CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	GTV-tongue GTV-prostate	Volume (cm <sup>3</sup> ) [Mean ± STD] DSC [Mean ± STD]	34, 28 ± 6 & 30 ± 7 87 ± 44 & 75 ± 22 0.83 ± 0.06 & 0.86 ± 0.06 0.68 ± 0.15 & 0.78 ± 0.07	iMAR improved the accuracy of the organ contouring
Axente et al. <sup>35</sup>	iMAR	Clinical (hip, H&N, spine, knee & breast); (n = 8)	Metal implants & metal prosthesis	CT <sub>art</sub> & CT <sub>cor</sub>	Selected structures	Visual conspicuity score [Median]	3 of 5 & 4 of 5	iMAR improved the accuracy and efficiency of contouring
Kovacs et al. <sup>54</sup>	iMAR	Clinical (H&N, spine & hip); (n = 1)	Metal implants	CT <sub>art</sub> , CT <sub>cor</sub> & CT <sub>ref</sub>	H&N Spine Hip	HD (mm) [Median] DSC of contours [Mean]	10.7, 5.1 & 2.9 18.2, 18.6 & 2.2 7.7, 3.5 & 2.0 0.75, 0.87 & 0.91 0.57 & 0.74 & 0.88 0.5 & 0.87 & 0.95	iMAR improved the contouring accuracy
Shirashi et al. <sup>55</sup>	SEMAR	Clinical (PelVis); (n = 40)	Iodine seeds with Ti & Ag	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Prostate gland	TPF [Mean ± STD]	0.982 ± 0.0159 & 0.992 ± 0.0103	SEMAR improved the accuracy of seed localization in brachytherapy

H&N Head and neck, Ti Titanium, Ag Silver, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, ICC Inter coefficient correlation, DSC Dice similarity coefficient, HD Hausdorff distance, TPF True-positive fraction, (w/o) without outliers and STD Standard deviation.

In another study by Andersson et al., visual grading was used for the evaluation of anatomical delineation.<sup>50</sup> A comparison between  $CT_{art}$  and either O-MAR or iMAR-corrected  $CT_{cor}$  was performed. Visual grading resulted in significantly ( $P < 0.001$ ) higher scores on both  $CT_{cor}$ . However, new induced artifacts were also identified on  $CT_{cor}$ . Kohan et al. used clinical H&N CT scans with dental implants for their O-MAR evaluation.<sup>51</sup> Reviewers with different experience levels in contouring performed area measurements of selected structures on  $CT_{art}$ ,  $CT_{cor}$  after O-MAR application, and  $CT_{ref}$ . The intra-class correlation coefficient (ICC) was calculated to assess inter-reader variability. The highest ICC value of 1 indicates the lowest inter-reader variability, while ICC value of 0 indicates the highest inter-reader variability. For all reviewers, the ICCs for  $CT_{art}$ ,  $CT_{cor}$  after O-MAR application, and  $CT_{ref}$  were 0.884, 0.971, and 0.989, respectively, without outliers; and 0.903, 0.948, and 0.985, respectively, with outliers. For the experienced readers, the ICC for  $CT_{art}$ ,  $CT_{cor}$  after O-MAR application, and  $CT_{ref}$  were 0.934, 0.975, and 0.990, respectively, without outliers; and 0.904, 0.979, and 0.976, respectively, with outliers. Application of O-MAR improved the ICC values and brought them closer to the reference ones. For this reason,  $CT_{cor}$  after O-MAR application reduced the inter-reader variability during contouring. Another study by Hansen et al. also concluded that the application of O-MAR on clinical H&N  $CT_{art}$  increased the organ delineation and contouring accuracy.<sup>52</sup> The authors also measured and compared the gross tumor volume (GTV) and parotid volume on  $CT_{cor}$  after O-MAR application and on  $CT_{art}$ . Removal of streak artifacts increased the signals from the artifact corrupted areas, thus depicting a consistently larger contoured GTV (mean 22%,  $P < 0.06$ ) and parotid volume (mean 7%,  $P = 0.05$ ). However, the authors noted that to determine the actual volumes of the delineated structures the measured volumes should be compared with the reference volumes. A similar finding was reported by Hagen et al.; they found that contouring on  $CT_{cor}$  after iMAR application would increase the mean GTV tongue in H&N cases.<sup>53</sup> The GTV tongue increased ( $P = 0.267$ ) from  $28 \pm 6 \text{ cm}^3$  (mean  $\pm$  STD) on  $CT_{art}$  to  $30 \pm 7 \text{ cm}^3$  (mean  $\pm$  STD) on  $CT_{cor}$ . However, the mean volume of the parotid as an OAR was reduced in this study. Moreover, the authors evaluated the size of the prostate GTVs on bilateral implanted pelvis CT scans, and they were reduced ( $P = 0.168$ ) from  $87 \pm 44 \text{ cm}^3$  (mean  $\pm$  STD) on  $CT_{art}$  to  $75 \pm 22 \text{ cm}^3$  (mean  $\pm$  STD) on  $CT_{cor}$  after iMAR application. For the OARs on  $CT_{cor}$  after iMAR application in the pelvis case, the mean volume for rectum and bladder was reduced and increased, respectively. The DSC of the contours with respect to the reference increased more for the  $CT_{cor}$  after iMAR application than for the  $CT_{art}$ . Both the GTV of the tongue and the prostate on  $CT_{cor}$  after iMAR application were underestimated in comparison with the predefined reference. Nevertheless, it improved the confidence in contouring, as indicated by higher DSC values. Axente et al. assessed the image quality and visual conspicuity of  $CT_{art}$  and  $CT_{cor}$  after iMAR application.<sup>35</sup> Different types of clinical images were used, such as hip cases with unilateral or bilateral metal implants, H&N cases with dental fillings, a spine with metal implants, a knee with prosthesis, and a breast with expander.

The median score for the image quality and visual conspicuity of  $CT_{art}$  and  $CT_{cor}$  after iMAR application increased from 3 to 4 of 5. During the image quality assessment, new secondary artifacts were identified on  $CT_{cor}$ . Another study investigated iMAR for its anatomical delineation accuracy.<sup>54</sup> For this study, clinical CT scans (H&N, spine, and hip) with metal implants were used. Maximum Hausdorff (HD) distance and DSC were calculated to quantify the anatomical delineation accuracy. High HD distance indicates large difference between two contours while short HD distance indicates small variation. Contouring was performed on  $CT_{art}$  and  $CT_{cor}$  after iMAR application and the maximum HD distance with respect to contours on  $CT_{ref}$  was 10.7 mm and 5.1 mm on dental scans, 18.2 mm and 18.6 mm on spine scans, and 7.7 mm and 3.5 mm on hip scans, for  $CT_{art}$  and  $CT_{cor}$ , respectively. In dental and hip scans, the maximum HD distance on  $CT_{art}$  had the largest values and differed significantly ( $P < 0.05$ ) from  $CT_{ref}$ . Furthermore, the calculated DSC values for contours on  $CT_{art}$  and  $CT_{cor}$  after iMAR application were 0.75 and 0.87 on dental, 0.57 and 0.74 on spine, and 0.5 and 0.87 on hip, respectively. The higher DSC values on  $CT_{cor}$  after iMAR application indicate that there is less variability in contouring and a higher accuracy.

The SEMAR algorithm was evaluated by Shiraishi et al. to quantify its ability to improve the detection accuracy of implanted iodine seeds which contain Ti and silver (Ag) in brachytherapy.<sup>55</sup> To identify seeds on both  $CT_{art}$  and  $CT_{cor}$  after SEMAR application an automatic seed finder was used, and the results were compared with reference positions. The mean true-positive fraction (TPF) was calculated, and it had significantly higher values ( $P < 0.05$ ) for  $CT_{cor}$  after SEMAR application ( $0.992 \pm 0.0103$ , [mean  $\pm$  STD]) than for  $CT_{art}$  ( $0.982 \pm 0.0159$ , [mean  $\pm$  STD]). Thus, the application of SEMAR on  $CT_{art}$  improved implanted seed detection.

Overall,  $CT_{cor}$  after application of the above-mentioned commercial MAR algorithms improved the anatomical conspicuity and contouring accuracy. On the other hand, new artifacts induced by the MAR algorithms appeared on resulting  $CT_{cor}$  and it is clear that the external factors such as physician's knowledge and experience considerably influence these results.

### 3.A.3 | Dose-based evaluations and corrections

#### Dosimetric impact of commercial MAR algorithms

The commercial MAR algorithms were also evaluated to assess their ability to improve the dose calculation accuracy in RT. Table 3 provides a summary of these research studies. O-MAR was evaluated by Kwon et al. in clinical and phantom studies.<sup>41</sup> Clinical H&N CT scans with dental implants were acquired in open-mouth and closed-mouth positions. In the close-mouth case, the nasopharynx, the parotid, and the submandibular salivary gland were contoured as targets and dose calculation was performed on  $CT_{art}$  and  $CT_{cor}$  after O-MAR application. A nonsignificant ( $P > 0.05$ ) mean gamma passing rate (% GP) of  $99.4 \pm 0.5\%$  (mean  $\pm$  STD) was reported between them using a 1%/1 mm gamma criterion. However, for targets such as the tongue and the tonsil, a large discrepancy in %GP was found between



**TABLE 3** Summary of the studies which investigated the dosimetric impact of commercial MAR algorithms in RT.

Author	MAR Algorithm	Images; No. of sample (n)	Metal	Type of radiation, Energy, and Mode of therapy	Dose calculation algorithm (TPS); [Dose measurement]	CT Scans	Calculation and Measurement		Results	Comments
Kwon et al. <sup>41</sup>	O-MAR	Phantom Clinical (H&N); [open mouth (n = 3) & closed mouth (n = 8)]	Al, Ti, Cr, & Zr Dental implants	Photon, 6 MV (VMAT)	AAA, (Eclipse™, Varian) [Gafchromic®, EBT2]	CT <sub>art</sub> & CT <sub>cor</sub>	Discrepancy in calculated dose (%) (3%/3mm) %GP (3%/3 mm) Mean %GP, on closed mouth (1%/1mm) %GP, on open mouth (1%/1mm)	95.3% (Ti) & 91.6% (Cr) 83.60% & 86.68% (Ti) 79.70% & 86.30% (Cr) 99.40 ± 0.5% 91.1%, 94.8% & 96.6%	Dose calculations after O-MAR application are closer to the measured dose	
Huang et al. <sup>56</sup>	O-MAR	Geometric phantom Anthropomorphic phantom	Ti & Cerrobend Ti & Amalgam	Photon, 6 MV (IMRT)	CCC (Pinnacle (Philips) [Gafchromic® EBT2 & ion chamber]	CT <sub>art</sub> & CT <sub>cor</sub> CT <sub>cor</sub> vs measurement CT <sub>art</sub> vs CT <sub>cor</sub>	Dose calculation error (%) (2%/2mm) Dose calculation error (%) (2%/2mm)	15% & 11.1% < 1% (Ti) 6.7% & 1.8% (Amalgam)	O-MAR had little dosimetric impact on Ti, and large impact on amalgam.	
Sillanpaa et al. <sup>49</sup>	O-MAR	Clinical (H&N); (n = 28)	Dental implants	Photons, 6 MV (IMRT & VMAT)	AAA, algorithm (Eclipse™, Varian)	(CT <sub>art</sub> - CT <sub>cor</sub> )/ ( CT <sub>cor</sub> )	Percent change Mean dose—PTV PTV V <sub>95%</sub>	-0.3% -0.2%	O-MAR algorithm did have an insignificant dosimetric effect	
Shen et al. <sup>57</sup>	O-MAR	Phantom Clinical (Spine); (n = 5)	Ti Spinal prosthesis	Photon, 6MV (IMRT) Photon, 6MV (IMRT)	CCC (Pinnacle, Philips) PB (Plan, BrainLAB Inc.)	CT <sub>art</sub> & CT <sub>cor</sub>	%GP (2%/2 mm) (1%/1 mm) %GP (2%/2 mm) (1%/1 mm)	> 99.98% > 99.96% 99.33% & 100% 96.96% & 99.93%	O-MAR did not affect the dose calculation accuracy for SBRT planning in patients with spinal implants.	
Akdeniz et al. <sup>59</sup>	O-MAR	Phantom	SS & Ti	Photons, 6 MV	AAA and AXB (Eclipse™, Varian) [Gafchromic™ EBT3]	(CT <sub>art</sub> vs measurement) & (CT <sub>cor</sub> vs measurement)	Dose differences at upper SS surface (%) in AXB in AAA Dose differences at lower Ti surface (%) in AAA	23.1% & 16.9% 18.2% & 12.7% -13.30% & -11.50%	Dose calculation after application of O-MAR with AXB shows closer dose agreement with dose measurements	
Jia et al. <sup>62</sup>	O-MAR	Phantom	SS	Proton	(Eclipse™, Varian)	CT <sub>cor</sub> & CT <sub>art</sub>	Relative dose discrepancy	20% at 2 mm & 80% at 4 mm	O-MAR improved the accuracy of dose distribution	

(Continues)

TABLE 3 (Continued)

Author	MAR Algorithm	Images; No. of sample (n)	Metal	Type of radiation, Energy, and (Mode of therapy)	Dose calculation algorithm (TPS); [Dose measurement]	CT Scans	Calculation and Measurement	Results	Comments
Righetto et al. <sup>53</sup>	O-MAR	Phantom	Ti	Proton	MC	(CT <sub>art</sub> vs reference) & (CT <sub>cor</sub> vs reference)	$\Delta$ WET (cm)	0.16 cm & 0.57 cm	Application of O-MAR overestimated the $\Delta$ WET
Andersson et al. <sup>50</sup>	O-MAR & iMAR	Phantom Clinical (H&N); (n = 30)	Dental filling Neck implant Dental implants	Proton, from 60 MeV to 226 MeV	Proton pencil beam (Eclipse™, Varian)	CT <sub>art</sub> & CT <sub>cor</sub> CT <sub>cor</sub> CT <sub>art</sub> & CT <sub>cor</sub>	$\Delta$ WET (mm) O-MAR iMAR Residual $\Delta$ WET O-MAR iMAR %GP (whole image)	-17.0 mm & -4.3 mm -16.1 mm & -2.3 mm -2.3mm -1.5 mm ~ 97%	O-MAR and iMAR improved the dose distribution. However, case-by-case evaluations before applying MAR is important
Bär et al. <sup>42</sup>	iMAR	Phantom Clinical (H&N); (n = 7) Clinical (Pelvis); (n = 1)	Al, Ti & SS Dental implant Hip bilateral implant	Photon, 6MV (IMRT)	CCC (RaySearch)	(CT <sub>cor</sub> & CT <sub>art</sub> ) vs CT <sub>ref</sub> CT <sub>art</sub> & CT <sub>cor</sub>	%GP (1%/1mm) Dose calculation error (%)	85.70% & 62.10% up to $\pm$ 5%	iMAR improved the dose calculation accuracy
Maerz et al. <sup>43</sup>	iMAR	Phantom	Dental implants	Photon (IMRT & VMAT)	PB and CC (Oncontra External Beam v4.3)	CT <sub>art</sub> & CT <sub>cor</sub>	%GP (3% / 3 mm)	90.6% & 96.2%	iMAR tends to improve the dose calculation accuracy
Axente et al. <sup>35</sup>	iMAR	Phantom	Al & SS	Photon, 6MV & 15MV (VMAT, 3D CRT) Proton, 195 MeV	AXB (Eclipse™, Varian) PBCS (Eclipse™, Varian)	CT <sub>art</sub> & CT <sub>cor</sub>	%GP for 6MV (2%/2mm) (2%/1mm) %GP, for 15MV (2%/2mm) (2%/1mm) %GP (2%/2mm) (2%/1mm)	97% & 99.4% 92.1% & 98.1% 97.5% & 99.6% 92.9% & 98.3% 88.6% & 91.9% 78.7% & 91%	iMAR improved the dose calculation accuracy in proton and photon beams
Huang & Kohli <sup>45</sup>	Smart MAR	Clinical (H&N); (n = 15) Clinical (Pelvis); (n = 10)	Dental filling Hip prosthesis	Photons, 6MV IMRT IMRT & VMAT	AAA (Eclipse™, Varian)	CT <sub>cor</sub> & CT <sub>art</sub>	Mean $\pm$ STD difference (%) CI D <sub>99%</sub> V <sub>100%</sub> CI D <sub>99%</sub> V <sub>100%</sub>	-0.3 $\pm$ 0.9% -0.1 $\pm$ 0.1% -0.1% $\pm$ 0.5% -8.8 $\pm$ 11.4% -0.1 $\pm$ 0.4% -8.8% 12.1%	Dosimetric differences between CT <sub>cor</sub> after Smart MAR and CT <sub>art</sub> were small

(Continues)

TABLE 3 (Continued)

Author	MAR Algorithm	Images; No. of sample (n)	Metal	Type of radiation, Energy, and (Mode of therapy)	Dose calculation algorithm (TPS); [Dose measurement]	CT Scans	Calculation and Measurement	Results	Comments
Inal & Sarpın <sup>46</sup>	Smart MAR	Phantom	Leads	Photons, 6MV (IMRT)	CS (Elekta XIO) [Matrix arrays]	(CT <sub>cor</sub> vs measurement) & (CT <sub>art</sub> vs measurement)	Mean %GP difference (3% /3 mm) 5 fields 7 fields 9 fields	94.98% & 96.11% 94.72% & 95.90% 91.34% & 92.83%	Smart MAR increased the gamma passing rate
Guilfoile et al. <sup>44</sup>	Smart MAR	Phantom	Hip prosthesis	Photon, 6MV, 10MV & 15 MV	Pinnacle (Philips) [ion chamber]	(CT <sub>cor</sub> vs measurement) & (CT <sub>art</sub> vs measurement)	Dose difference (%)	Refer to fig. 6 in Ref. [44]	Smart MAR improved the dose calculation accuracy in treatment planning
Murazaki et al. <sup>47</sup>	SEMAR	Phantom	Unilateral & bilateral metal	Photons, 10MV (one filed irradiation, opposite portal irradiation, four field irradiation, & VMAT)	AAA and AXB (Eclipse™, Varian)	CT <sub>ref</sub> vs CT <sub>art</sub> CT <sub>ref</sub> vs CT <sub>cor</sub>	%GP (1%/0 mm) in AXB with opposite portal irradiation	89.89% 95.03%	Application of SEMAR improved the accuracy of dosimetry in treatment planning
Miki et al. <sup>48</sup>	SEMAR	Phantom Clinical (H&N); (n = 2)	Metal screws Dental implants	Carbon ion	N/A	CT <sub>art</sub> & CT <sub>cor</sub>	Mean %GP (3%/ 3mm) (2%/2mm) PTV-D95% %GP (3%/3mm) Clinical 1 Clinical 2 PTV-D95% Clinical 1 Clinical 2	89.1% 79% 82.4% & 95.4% 99.9% 92.7% 92.2% & 92.5% 90.9% & 95.7%	Application of SEMAR improved the dose distribution
Shiraishi et al. <sup>55</sup>	SEMAR	Clinical (pelvis); (n = 40)	Iodine capsules with Ti & Ag	Brachytherapy	VariSeed 8.0 (Varian Medical Systems)	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Median difference D <sub>90%</sub> V <sub>100%</sub> V <sub>150%</sub>	0.78% & 0.28% 0.24% & 0.06% 1.70% & 0.58%	Prostate postimplant CT <sub>cor</sub> after SEMAR improved the postimplant dosimetric parameters

H&N Head and neck, Ti Titanium, SS Stainless Steel, Ag Silver, Cr Chromium, Zr Zirconia, IMRT Intensity-Modulated Radiation Therapy, VMAT Volumetric-Modulated Arch Therapy, CCC Convolution Collapsed Cone, PB Pencil Beam, AAA Anisotropic Analytical Algorithm, AXB Acuros®XB, CS Convolution Superposition, MC Monte-Carlo, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, vs versus, %GP gamma passing rate, ΔWET Deviation in Water Equivalent Thickness, PTV Planning Target Volume.

CT<sub>art</sub> and CT<sub>cor</sub> after O-MAR application under the same criteria in the open-mouth scans. A virtual build-up region was created in front of the oral cavity due to the metal artifacts on CT<sub>art</sub> which resulted in this dose discrepancy. Furthermore, the phantom studies revealed that the calculated dose on CT<sub>cor</sub> after O-MAR application was closer to the delivered dose (measured with films) than the calculated dose on CT<sub>art</sub>. Also Huang et al. studied the application of O-MAR to improve dose calculation using phantoms.<sup>56</sup> A custom-made geometric phantom with inserts of Ti or Cerrobend, and two anthropomorphic phantoms, one with spinal implants and another one with dental fillings, were used. In the geometric phantom with Ti insert, the dose errors between calculated and measured dose with a 2%/2 mm gamma criterion were 15% for CT<sub>art</sub> and 11.1% for CT<sub>cor</sub> after O-MAR application in the region which was under the metal implant. Furthermore, in the anthropomorphic phantoms' evaluation, the O-MAR application improved the dose calculation accuracy in the dental filling case while it had little impact for the spinal implant case. Similar findings were reported by Sillanpaa et al. for clinical H&N CT scans.<sup>49</sup>

The dosimetric impact of O-MAR on the spine for stereotactic body radiation therapy (SBRT) was evaluated by Shen et al. using a phantom and clinical CT scans.<sup>57</sup> SBRT can deliver high ablative radiation dose to the target, while sparing OARs.<sup>58</sup> A CT electron density phantom (Gammex, Model RMI 465, Middleton, WI, USA) with a Ti implant was used in this study. A similar calculated planar dose distribution was observed on CT<sub>art</sub> and CT<sub>cor</sub> after O-MAR application and %GP was larger than 99.98% for a 2%/2 mm gamma criterion and 99.96% for a 1%/1 mm gamma criterion. The evaluation of the clinical CT scans revealed similar findings. Thus, the study concluded that O-MAR does not significantly affect the dose calculation accuracy and can, therefore, be safely used for SBRT treatment planning. In another recent study, Akdeniz et al. evaluated the dosimetric effects of metal implants in small-field RT using a custom-made slab phantom (PTW, Freiburg, Germany).<sup>59</sup> The study revealed that small differences in dose were observed between the calculated and measured dose on both CT<sub>art</sub> and CT<sub>cor</sub> after O-MAR application. The authors found that the type of dose calculation algorithms available on the TPS also influences the dose differences. The Anisotropic Analytical Algorithm (AAA)<sup>60</sup> and Acuros External Beam (AXB) algorithm<sup>61</sup> available in the Eclipse™ TPS (Varian Medical Systems, Palo Alto, CA) were evaluated. On CT<sub>cor</sub> after O-MAR application the use of the AXB algorithm better reduced the dose differences between the calculated and measured dose compared to the AAA algorithm.

Jia et al. studied the dosimetric impact of O-MAR on proton therapy treatment planning using a solid water phantom with a SS crew inserted.<sup>62</sup> The discrepancies of relative depth dose distribution were calculated on CT<sub>cor</sub> after O-MAR application and CT<sub>art</sub>, and they were 2 mm at 20% relative dose and 4 mm at 80% relative dose, respectively. The O-MAR and iMAR algorithms were evaluated for proton therapy in phantom and clinical studies by Andersson et al.<sup>50</sup> A head phantom (CIRS, Model 731-HN, Norfolk, VA, USA) was CT scanned with and without a removable dental filling and a neck implant. Deviation in water equivalent thickness ( $\Delta$ WET) was

calculated to find the proton range errors on CT<sub>ref</sub>, CT<sub>art</sub>, and CT<sub>cor</sub>. In case of the dental filling, along a dark streak,  $\Delta$ WET was improved from  $-17.0$  mm to  $-4.3$  mm by O-MAR, and from  $-16.1$  mm to  $-2.3$  mm by iMAR. For other directions,  $\Delta$ WET increased or remained unchanged on CT<sub>cor</sub>. Generally,  $\Delta$ WET was reduced in case of the neck implant; however, residual deviations up to  $-2.3$  mm with O-MAR and up to  $-1.5$  mm with iMAR remained. In the clinical H&N study with dental implants, planned dose distributions to a neck node were calculated on CT<sub>art</sub> and CT<sub>cor</sub>, and minor differences were observed. On the other hand, in a phantom (Gammex Inc., Middleton, WI, USA) study with a Ti implant, Righetto et al. reported that the calculated proton range using WET was overestimated more on the CT<sub>cor</sub> after O-MAR application than on both CT<sub>art</sub> and CT<sub>ref</sub>.<sup>63</sup> The reference was obtained using stopping power values from the data which were published by NIST (Gaithersburg, MD, USA). As a result,  $\Delta$ WET on CT<sub>cor</sub> after O-MAR application and CT<sub>art</sub> were 0.57 cm and 0.16 cm, respectively, when compared with the reference value.

Bär et al. focused their work on the iMAR algorithm using CT scans from the Gammex 467 phantom with and without the metal inserts, and the clinical CT scans of H&N and a hip with implants.<sup>42</sup> In the phantom, a target was contoured in the center between two implanted metals, and dose calculations were performed on CT<sub>ref</sub>, CT<sub>art</sub>, and CT<sub>cor</sub> after iMAR application. A %GP for a 1%/1 mm gamma criterion of 62.1% was found between the CT<sub>ref</sub> and CT<sub>art</sub> while CT<sub>cor</sub> improved it up to 85.7%. In the clinical CT scans, dose differences up to  $\pm 5\%$  have been shown when comparing the plans which were calculated on CT<sub>cor</sub> and CT<sub>art</sub>. The application of iMAR reduced the dose errors and, therefore, the authors recommended to use this method for RT applications. A similar finding was reported by another cylindrical phantom study with dental implants. The dose distributions were calculated on CT<sub>art</sub> and CT<sub>cor</sub> and it was shown that the application of iMAR improved the %GP from 90.6% to 96.2% for a 3%/3 mm gamma criterion.<sup>43</sup> In the CIRS phantom study,<sup>35</sup> the calculated dose distribution on CT<sub>cor</sub> after iMAR application was closer to the CT<sub>ref</sub> dose distribution than the dose distribution calculated on CT<sub>art</sub>. In addition to dose distributions, the %GP of photons (6 MV and 15 MV) and protons (195 MeV) were derived from the calculated dose distributions on CT<sub>art</sub> and CT<sub>cor</sub>. For the 6MV photon beam, iMAR increased the photon %GP for a 2%/2 mm gamma criterion from 97% to 99.4% and for the 15 MV photons similar results were shown. For the proton beam, iMAR increased the %GP from 88.6% to 91.9% under the same criteria.

Smart MAR was evaluated for dose calculation accuracy in RT by Huang and Kohli using clinical CT scans of H&N with dental fillings and of a pelvis with a metal prosthesis.<sup>45</sup> The average conformity index (CI),  $D_{99\%}$ , and  $V_{100\%}$  were calculated on CT<sub>cor</sub> after Smart MAR application and CT<sub>art</sub> and then compared. The average percentage (mean  $\pm$  STD) differences in CI,  $D_{99\%}$ , and  $V_{100\%}$  on H&N CT scans were  $-0.3\% \pm 0.9\%$ ,  $-0.1\% \pm 0.1\%$ , and  $-0.1\% \pm 0.5\%$ , respectively. For the CT scans of the pelvis, they were (mean  $\pm$  STD)  $-8.8\% \pm 11.4\%$ ,  $-0.1\% \pm 0.4\%$ , and  $-8.8\% \pm 12.1\%$ , respectively. Also, this study found that the calculated dose differences between

the  $CT_{cor}$  and  $CT_{art}$  were not significant. In another study, Inal and Sarpün evaluated Smart MAR for dose calculation accuracy in 12 different intensity-modulated radiation therapy (IMRT)<sup>64</sup> plans with 5-, 7-, and 9-field beam arrangements and segment numbers.<sup>46</sup> IMRT is an inverse planning-based treatment delivery method which is optimized until the target volume and normal tissue reached the prescribed dose. A custom-made metal-containing phantom was CT scanned. For the five, seven, and nine fields, the calculated dose distributions on  $CT_{art}$  and  $CT_{cor}$  after Smart MAR application were compared with the measured dose. The %GP for a 3%/3 mm gamma criterion of the above-mentioned dose distribution were 94.98% and 96.11% (for five fields), 94.72% and 95.90% (for seven fields), and 91.34% and 92.83% (for nine fields), respectively. The improvement of %GP on  $CT_{cor}$  shows that Smart MAR increased the dose calculation accuracy. A similar finding was reported by Guilfoile et al. using a custom-made hip phantom with bilateral hip prostheses.<sup>44</sup>

Murazaki et al. evaluated the SEMAR algorithm for dose calculation accuracy using a standard electron density phantom (CIRS, Model 062A, Norfolk, VA, USA) with metal inserts.<sup>47</sup> Several different treatment plans were prepared, using forward planning with one field, with two opposite fields, and four fields; and with volumetric-modulated arc therapy (VMAT).<sup>65</sup> The AXB and AAA algorithms were used to calculate the dose distribution on  $CT_{art}$  and  $CT_{cor}$  after SEMAR application. In AXB for the two opposite fields, SEMAR increased the %GP for a 1%/0 mm criterion from 89.89% to 95.03%, and similar results were observed also in AAA as well as for the other forward planning methods under the same criteria. This implies that the calculated dose distribution was improved with the use of SEMAR. Miki et al. studied the dosimetric impact of the SEMAR algorithm in carbon ion therapy using phantom and clinical CT scans.<sup>48</sup> An anthropomorphic head phantom was scanned with and without metal inserts and then SEMAR was applied to create  $CT_{cor}$  from the  $CT_{art}$ . A planning target volume (PTV)<sup>66</sup> was placed on the dark streak band and PTV- $D_{95\%}$  was measured. The calculated PTV- $D_{95\%}$  on clinical H&N scans with dental implants and with tumors near the spinal cord as the treatment target showed a higher PTV- $D_{95\%}$  (from 82.4% to 95.4%) for  $CT_{cor}$  than for  $CT_{art}$ . To investigate brachytherapy applications, Shiraiishi et al. used clinical pelvis CT scans with implanted iodine seeds (which contain Ti and Ag).<sup>55</sup> SEMAR was applied to generate  $CT_{cor}$  from  $CT_{art}$ , and then  $D_{90\%}$ ,  $V_{100\%}$ , and  $V_{150\%}$  were calculated on those CT scans and compared with  $CT_{ref}$ . The differences in dosimetric calculations were significantly smaller ( $P < 0.05$ ) between  $CT_{cor}$  and  $CT_{ref}$  than between  $CT_{art}$  and  $CT_{ref}$ .

In general, the  $CT_{cor}$  after application of a commercial MAR algorithm improves the dose calculation accuracy in RT. This improvement is indicated by increment of %GP, reduction of errors in dose calculation compared to  $CT_{art}$ , and minor differences from the measured dose using films and/or ion chambers or calculated dose on  $CT_{ref}$ . Furthermore, improvements in the WET calculation were also reported after artifact correction by commercial MAR algorithms in proton therapy. However, the amount of improvement in calculated dose depends on the utilized dose calculation algorithm and on

radiation therapy technique. Studies reported that the AXB algorithm increases the dose calculation accuracy more than the AAA. Moreover, application of VMAT is preferable to improve the accuracy in calculated dose compared to IMRT.

### Dosimetric impacts of density correction methods

The density correction methods which are available in the TPSs can also be used to reduce the metal artifacts on CT scans for RT applications, see Table 4. Maerz et al. used a dental implant cylindrical phantom to evaluate dose calculation accuracy after application of a density correction method.<sup>38</sup> IMRT and VMAT plans were calculated on  $CT_{art}$  and  $CT_{cor}$  after density correction and then compared with dose measurements using films. Findings revealed that the accuracy of dose calculation was higher ( $P = 0.015$ ) on  $CT_{cor}$  than on  $CT_{art}$  for both IMRT and VMAT plans. Moreover, the VMAT plan increased the %GP for a 3%/3mm gamma criterion in comparison with the IMRT plan on both  $CT_{art}$  and  $CT_{cor}$ . On the other hand, Acquah et al. used a phantom (CIRS, Model 002LFC, Norfolk, VA, USA) to compare the calculated dose on  $CT_{cor}$  after density correction with the measured dose using ion chambers.<sup>67</sup> The authors found that treatment planning on  $CT_{cor}$  gave a 16% higher average dose discrepancy. The dose discrepancy between the calculated dose on  $CT_{art}$  and the measured one was not mentioned in this study. This study suggested that caution should be taken while planning on  $CT_{cor}$  after density correction. Parenica et al. evaluated the impact of a density correction method on a VMAT plan using CT scans of a custom-made phantom and on clinical pelvis CT scans with hip prostheses.<sup>68</sup> Densities of the prostheses and surrounding tissue which contain metal artifacts were overridden with the appropriate corrected density. Dose calculation algorithms with collapsed cone convolution superposition (CCCS)<sup>69</sup> available on Pinnacle (Philips, Fitchburg, Wisconsin) and a Monte Carlo<sup>70</sup>-based algorithm available in Monaco (ELEKTA, Stockholm, Sweden) TPS were used. The calculated dose on  $CT_{art}$  and  $CT_{cor}$  after density override was compared with the respective measured dose using a thermoluminescence dosimeter (TLD). In the phantom, the dose errors were 9.2% and 4.4% for Pinnacle and 3.6% and 0.2% for Monaco, respectively. For the clinical pelvis scans, the prostate was contoured as the target. Subsequently, a comparison of the calculated average  $D_{95\%}$  without and with density correction was 99.3% and 82.7% for Pinnacle and 99.0% and 90.6% for Monaco, respectively. The reduced  $D_{95\%}$  on  $CT_{cor}$  resulted from a reduced PTV. The same authors in another study<sup>71</sup> found similar results for density correction method on data of a phantom with Al, Zn, and SS inserts, and clinical H&N CT scans with dental implants.

## 3.B | Research MAR methods

### 3.B.1 | Traditional MAR algorithms

Several research MAR algorithms which utilize traditional image processing methods have been proposed and/or evaluated for the reduction in metal artifacts on CT scans for RT applications. These algorithms rely on several different principles. Some approaches are

**TABLE 4** Summary of the studies which investigated the dosimetric impact of density correction in RT.

Author	Images; No. of sample (n)	Metal	Beam, Energy, and (Mode of therapy)	Dose calculation algorithm (TPS); [Dose measurement]	CT Scans	Calculation and Measurement	Results	Comments
Maerz et al. <sup>38</sup>	Phantom	Dental implants	Photon, 6 MV (VMAT & IMRT)	PB and CC (Oncentra External Beam)	CT <sub>art</sub> & CT <sub>cor</sub>	%GP IMRT VMAT	0.954% & 0.980% 0.983% & 0.990%	Density correction increased the dose calculation accuracy
Acquah et al. <sup>67</sup>	Phantom	Spine implant	Photons, 6 MV & 15MV (3D-CRT)	CC & PB (Oncentra External Beam)	CT <sub>cor</sub> vs CT <sub>art</sub>	Mean dose discrepancy (%)	16%	Density correction resulted in higher dose discrepancy
Paronica et al. <sup>68</sup>	Phantom Clinical (Pelvis); (n = 6)	Ti Hip prosthesis	Photons, 6 MV (VMAT)	CCC (Pinnacle, Philips) & MC (Monaco) [Ion chamber]	(CT <sub>art</sub> & CT <sub>cor</sub> ) vs Measured CT <sub>art</sub> & CT <sub>cor</sub>	Dose errors (%) Pinnacle Monaco Average D <sub>95%</sub> Pinnacle Monaco	9.2% & 4.4 % 3.6% & 0.2% 99.3% & 82.7% 99.0% & 90.6%	Density override can be used to optimize the dose calculation planning with MC
Paronica et al. <sup>71</sup>	Phantom Clinical (H&N); (n = 9)	AL, Zn & SS Dental implants	Photons, 6 MV (VMAT)	CCC (Pinnacle, Philips) & MC (Monaco) [TLD]	(CT <sub>art</sub> & CT <sub>cor</sub> ) vs Measured CT <sub>art</sub> & CT <sub>cor</sub>	Dose errors (%) for Al Pinnacle Monaco for Zn Pinnacle Monaco for SS Pinnacle Monaco D <sub>95%</sub>	4% & 4% 3.9% & 1.7% 8.6% and 7.1% 9.8% and 2.3% 8.8% and 5.8% 9.5% and 3.0% Refer to Table 5 in Ref. [71]	The density override reduced the potential risk of compromising the dose to the target and healthy tissues

Ti Titanium, SS Stainless Steel, Ag Silver, Zn Zinc, CRT Conformal Radiation Therapy, IMRT Intensity-Modulated Radiation Therapy, VMAT Volumetric-Modulated Arch Therapy, CCC Convolution Collapsed Cone, PB Pencil Beam, MC Monte-Carlo, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, vs versus, %GP gamma passing rate.

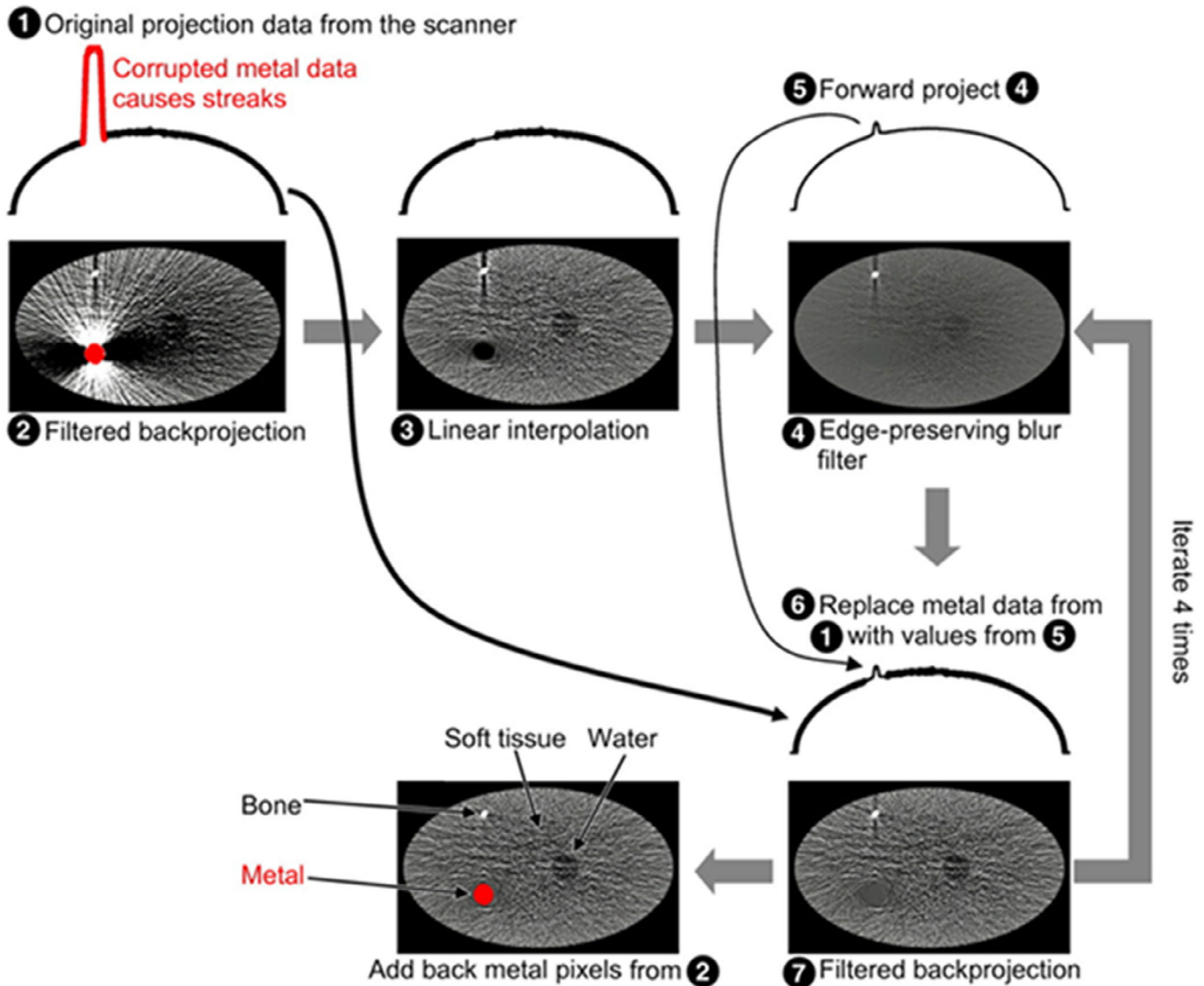
based on image inpainting or sinogram inpainting<sup>72–75</sup>; some others require the acquisition of additional tilted CT scans,<sup>76</sup> propose novel image acquisition, and reconstruction methods<sup>77,78</sup>; or require the use of magnetic resonance imaging (MRI)<sup>79,80</sup> or megavoltage CT (MVCT).<sup>81,82</sup>

For the metal deletion technique (MDT)<sup>75</sup> (Fig. 4) which uses sinogram inpainting iteratively, initially pixels which contain metal data are segmented from CT<sub>art</sub> (Fig. 4, image 2). Then, linear interpolation (LI)<sup>83</sup> and edge-preserving blur filters (Fig. 4, image 4) are applied on this CT scan to calculate the missing pixel values and to reduce the noise, respectively. Subsequently, the linearly interpolated and noise-reduced image is forward projected to create an initial sinogram (Fig. 4, number 5). This sinogram is used iteratively (four iterations in total) to replace the pixels which contain metal artifacts in the original sinogram. On each iteration, rays that pass through the metal are replaced with the value from the previous iteration (Fig. 4, number 6). This procedure results in a corrected sinogram. Finally, the filtered back-projection of the corrected sinogram (Fig. 4, image 7) with added metal data produces the CT<sub>cor</sub>.

An MRI-based MAR algorithm was proposed by Park et al.<sup>79</sup> The proposed method reduces the metal artifacts by mapping the HU values from a nearby artifact-free CT slice using a coregistered MRI scan. Initially, the CT<sub>art</sub> slice and adjacent artifact-free slice are

manually identified and then registered with their corresponding MRI slice. Based on the intensity values of pixels on the paired MRI slice and HU values from the artifact-free CT slice, HU values of pixels on CT<sub>art</sub> slice are calculated. Nielsen et al. evaluated their proposed MR-based MAR algorithm (kerMAR)<sup>80</sup> to reduce the metal artifacts.<sup>84</sup> kerMAR requires aligned CT and MRI scans of the same anatomy. It uses a Bayesian modelling<sup>85</sup> approach to compute the corrected HU values of a corrupted CT slice from the corresponding coregistered MRI slice.

Liugang et al. proposed a MAR method which makes use of megavoltage computed tomography (MVCBCT) and kilovoltage computed tomography (kVCT) scans for MAR.<sup>81</sup> Initially, the metal part in the MVCBCT scan is segmented and forward projected to obtain the metal trace. Then, a prior image is obtained by combining both scans through the fusion method proposed by Wang et al.<sup>86</sup> and this image is forward projected to obtain missing data to replace the metal trace on the kVCT scan. Finally, the kVCT scan with reduced metal artifacts is created through filtered back projection. In a similar study by Jeon et al., a hybrid sinogram-based MAR method was proposed for helical tomotherapy.<sup>82</sup> It also requires a kVCT scan and a MVCT scan of the same anatomy. During the MAR application, it replaces the metal affected signals of the forward-projected kV sinogram with the corresponding forward-projected MV sinogram.



**FIG. 4.** Working principle of the MDT MAR algorithm. A metal segmented CT slice (image 2) undergoes linear interpolation (image 3) and edge preserving filtration (image 4). Then, the filtered image is forward projected (number 5) and used to replace the metal data on the original projection (number 1). The resulting projection (number 6) is filtered back projected and undergoes edge preserving filtering. This process is iterated four times and results in a metal artifact reduced CT slice (image 7). Finally, the segmented metal is added to create the artifact-corrected CT slice. (Reproduced, with permission, from Boas F E, Fleischmann D. Evaluation of Two Iterative Techniques for Reducing Metal Artifacts in Computed Tomography. *Radiology* 2011; 259:894-902. The Radiological Society of North America (©RSNA)).

Finally, the application of filtered back projection on the corrected kVCT sinogram produces a kVCT scan with corrected artifacts.

A study by Kim et al. proposed to acquire an additional tilted CT scan in which less metal artifacts are present.<sup>76</sup> First, an artifact map is generated from the denoised initial CT scan ( $CT_{art}$ ) and the additional tilted CT scan. Then, utilizing the SSIM metric, correlation maps are generated from the  $CT_{art}$  and the additional CT scan in comparison with the artifact map. A minimum value in the correlation map represents smaller amounts of metal artifacts and, thus, the  $CT_{cor}$  is reconstructed using the minimum correlation values.

A new CT reconstruction algorithm, augmented likelihood image reconstruction (ALIR)<sup>78</sup> was investigated for MAR for RT applications. It utilizes an iterative scheme and prior information of the metal implant such as shape and attenuation coefficient. A new

sinogram is created by masking the metal on the initial sinogram. Using this new sinogram an image is iteratively reconstructed using log-likelihood function<sup>87</sup> and prior information. During each iteration, a bilateral filter is applied on the reconstructed image to reduce the remaining metal artifacts and then a forward projection of this images is used to calculate new projection values, which are combined with the initial projection values thereafter. Finally, the CT scan is reconstructed with reduced metal artifacts.

Dong et al. developed an image-based MAR method which utilizes the anatomical similarity in adjacent CT slices which are free from the metal artifacts ( $CT_{ref}$ ).<sup>72</sup> Initially, a gamma map is calculated which is a weighted summation of the relative HU error and the distance error for each pixel between the  $CT_{art}$  slice and the relative artifact-free CT slice. Then, the minimum value in the

gamma map is used to identify and replace the artifact-corrupted pixels on the  $CT_{art}$  slice by the corresponding pixels on the artifact-free CT slice.

An MAR method which includes a sinogram precorrection (which is not described) and a hardware adaptation for proton therapy was proposed by Jin et al.<sup>77</sup> The hardware adaptation includes an increase in the X-ray energy from 120 kVp (standard energy) to 180kVp, and an increase in the triggering rate of data acquisition system (DAS)<sup>88</sup> of the X-ray detectors. In the end, the prelog iterative CT reconstruction method<sup>89</sup> is applied to yield the resulting  $CT_{cor}$ .

Miksys et al. studied image-based MAR methods and a sinogram-based MAR method to reduce the metal artifacts from and around implanted radioactive seeds which contain Ag and Ti for brachytherapy procedures.<sup>74</sup> The image-based MAR methods include simple threshold replacement (STR), 3D median filter, and the generation of a virtual sinogram. STR utilizes a HU thresholding technique to identify the seeds and bright spot artifacts,<sup>90</sup> and correct them by replacing them with predefined HU value. The 3D median filter instead reduces the metal artifacts by smoothing the high voxel values which are identified through HU thresholding. In the virtual sinogram generation, first the metal on  $CT_{art}$  is segmented using thresholding and then  $CT_{art}$  and the segmented metal are forward projected. Subtraction of the metal projection from the  $CT_{art}$  projection resulted in a partial sinogram. Finally, through the application of interpolation and filtered back projection, the partial sinogram yields the  $CT_{cor}$ . The sinogram-based MAR method also utilizes similar steps, but instead of generating a virtual sinogram, it uses the sinogram from the CT scanner.

### Evaluation of HU values retrieval and image quality

Research-based MAR algorithms based on traditional image processing methods were evaluated for their ability to restore HU values and their ability to improve image quality of CT scans which will be used for RT applications. Table 5 summarizes the details of the corresponding publications.

The proposed MRI-based MAR algorithm by Park et al. was evaluated for its ability to improve the HU values.<sup>79</sup> Simulated metal artifacts were introduced on clinical H&N and brain  $CT_{ref}$ , and then the proposed MAR algorithm was applied to generate a  $CT_{cor}$ . The comparison of  $CT_{art}$  and  $CT_{cor}$  with  $CT_{ref}$  showed a reduction in absolute mean HU difference from 495 HU to 108 HU in the brain, and from 370 HU to 92 HU in the H&N. The HU of  $CT_{ref}$  was not reported. Nielsen et al. evaluated their kerMAR method for image quality improvements.<sup>84</sup> During the image quality evaluation, CT scans from a custom-made veal shank phantom with metal inserts and clinical H&N CT scans with dental implants were used. In addition to kerMAR, also O-MAR was applied to the CT scans for a comparison. On the scans of the phantom, kerMAR and O-MAR reduced the metal artifacts similarly ( $P > 0.05$ ); however, on the clinical CT scans kerMAR outperformed O-MAR ( $P < 0.01$ ). This difference can be explained by the fact that O-MAR produces residual artifacts and by the fact that kerMAR uses corresponding MRI images as an

external source of prior information. However, if the acquired CT and MRI scans are not aligned properly, the kerMAR method introduces new artifacts as well.

The MVCBCT- and kVCT-based MAR method<sup>81</sup> was evaluated for HU value improvement using an intensity-modulated verification phantom (CIRS, Norfolk, VA, USA) with SS inserts and clinical H&N CT scans with dental implants. The study found that the proposed MAR method reprojected the HU values with high accuracy. The hybrid sinogram-based MAR method by Jeon et al. was evaluated using a standard “cheese” phantom (Standard Imaging, Middleton, USA) with metal inserts made of Ag, tin (Sn), copper (Cu), and iron (Fe), and clinical H&N CT scans with dental implants.<sup>82</sup> Application of the proposed method improved the accuracy of true densities and HU values for metal and non-metal tissues (e.g., air, lung, and bone) on the phantom and the clinical scans, respectively. In addition to the HU value improvements, also CNR and SNR improved on the  $CT_{cor}$  after application of the hybrid sinogram-based MAR in comparison with  $CT_{art}$ .

The MAR method which requires the acquisition of an additional tilted CT scan<sup>76</sup> was evaluated for its ability to improve HU values. In an experimental study, a customized RANDO head phantom (The Phantom Laboratory Inc., NY, USA) with dental implants was used, while in a simulation study an XCAT numerical phantom<sup>91</sup> with bilateral hip implants was used. The mean absolute percentage errors in HU values were calculated for  $CT_{art}$  and  $CT_{cor}$  in comparison with  $CT_{ref}$ . This showed a reduction for  $CT_{cor}$  of up to 89% in the experimental study, and up to 86% in the simulation study. Dong et al. evaluated their image-based MAR method on clinical H&N CT scans with dental fillings and on a CT scan of a pelvis with simulated hip implants.<sup>72</sup> The mean HU error was calculated on both  $CT_{art}$  and  $CT_{cor}$  compared with  $CT_{ref}$  and it reduced from 360 HU to 24 HU in the H&N case, and from 460 HU to 34 HU in the pelvis case. Thus, the study concluded that the proposed image-based MAR can be used to reduce the metal artifacts and improve HU for RT applications.

The proposed image-based MAR methods and sinogram-based MAR method for brachytherapy by Miksys et al. were applied on CT scans of a custom-made agarose gel phantom and on clinical prostate CT scans with implanted seeds.<sup>74</sup> The STR removed bright spot artifacts around the seeds, but failed to mitigate any dark streak artifacts. Although the 3D median filter successfully mitigated the white and dark streak artifacts, it blurred the resulting  $CT_{cor}$ . The virtual sinogram and the raw sinogram methods reduced the streak artifacts, however, they introduced new artifacts on  $CT_{cor}$ . The authors concluded that the image-based MAR methods, especially the STR and 3D median filter, were more successful in reducing the artifacts in comparison with the sinogram-based MAR.

### Dosimetric impacts of traditional MAR algorithms

In addition to the improvement of HU and image quality, dosimetric impacts of research-based traditional MAR algorithms were evaluated in several research studies. Table 6 describes the findings of these evaluations.



**TABLE 5** Summary of the studies which investigated HU restoration and image quality improvement of research-based MAR algorithms for RT application which were based on traditional image processing methods.

Author	MAR	Images; No. of sample (n)	Metals	CT scans	ROI; [Reference HU value (Mean)]	Measurements	Results	Findings
Park et al. <sup>79</sup>	MRI-based CT MAR	Clinical (H&N) & brain); (n = 2)	Simulated artifacts	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Whole CT scan Brain; [Not reported] H&N; [Not reported]	Absolute mean HU different	495 HU & 108 HU 370 HU & 92 HU	MRI-based CT MAR improved the HU value measurement
Nielsen et al. <sup>84</sup>	kerMAR	Phantom Clinical (H&N); (n = 9)	Metal pins Dental implants	CT <sub>art</sub> vs CT <sub>cor</sub> (kerMAR) + CT <sub>cor</sub> (O-MAR) CT <sub>cor</sub> (O-MAR) vs CT <sub>cor</sub> (kerMAR)	Soft tissue Bone Oral cavity Mandible Teeth	Differences in low/high frequencies in HU	Refer to figs. 6 & 7 in Ref. [84] Refer to fig. 6 & 7 in Ref. [84]	kerMAR can be used for restoration of the HU values
Liugang et al. <sup>81</sup>	MVCBCT & kVCT method	Phantom Clinical (H&N); (n = 1)	SS with different diameter Dental implant	CT <sub>art</sub> & CT <sub>cor</sub> & CT <sub>ref</sub> (CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> & CT <sub>ref</sub> )	Pixel value along horizontal & vertical lines on CT scan Whole CT Scan	HU Percentage HU difference	Refer to fig. 8 in Ref. [81] Refer to fig. 7 in Ref. [81]	Proposed MVCBCT & kVCT methods improved the accuracy of HU value measurements
Jeon et al. <sup>82</sup>	Hybrid sinogram-based MAR	Phantom Clinical (H&N); (n = 3)	Ag, Sn, Cu & Fe Dental filling	CT <sub>art</sub> (kVCT) vs CT <sub>ref</sub> , CT <sub>art</sub> (MVCT) vs CT <sub>ref</sub> & CT <sub>cor</sub> vs CT <sub>ref</sub> CT <sub>art</sub> (MVCT) vs CT <sub>ref</sub> & CT <sub>cor</sub> vs CT <sub>ref</sub>	Non-metal tissue; [Not reported] Metal inserts; [Not reported] Whole image; [Not reported]	Mean true density errors Visual quality	1.81%, 0.42% & 1.94% 0.48% & 0.30% Refer to fig. 9 in Ref. [82]	Hybrid sinogram-based MAR can be used for image quality and HU values improvement
Kim et al. <sup>76</sup>	Additional tilted CT scan-based MAR	Simulation Phantom Experimental phantom	Bilateral hip implant Dental implants	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Whole image; [Not reported]	Mean HU absolute percentage error (%)	62.98% & 8.52% 94.12 & 10.12%	Metal artifacts can be reduced by the proposed additional tilted CT scan-based MAR
Dong et al. <sup>72</sup>	Image-based MAR	Clinical (Hip); (n = 1) Clinical (H&N); (n = 1)	Simulated artifacts Dental fillings	CT <sub>art</sub> & CT <sub>cor</sub>	Whole image; [Not reported]	Mean HU error	360 HU & 24HU 460 HU & 34 HU	Image-based MAR improved the accuracy of measured HU values

H&N Head & Neck, Ti Titanium, SS Stainless Steel, Ag Silver, Sn Tin, Cu Copper, Fe Iron, kVCT kilovoltage CT, MVCT Megavoltage CT, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, vs versus.

**TABLE 6** Summary of the studies which investigated the dosimetric impact in RT of research-based MAR algorithms based on traditional image processing.

Author	MAR	Images; No. of Samples (n)	Metal	Beam, Energy, and (Mode of therapy)	Dose calculation algorithm (TPS); [Dose measurement]	CT Scans	Calculation and Measurement	Results	Comments
Park et al. <sup>79</sup>	MRI-based CT MAR	Clinical (H&N); & brain); (n = 2)	Simulated artifacts	Proton	N/A	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> & CT <sub>ref</sub> )	Absolute mean ΔWET	1.7cm & 2mm 2.4cm & 2mm	The proposed method increased the proton range calculation accuracy
Nielsen et al. <sup>84</sup>	kerMAR	Phantom Clinical (H&N); (n = 9)	Metal pins Dental implants	Photon (6 MV) Electron (12 MeV) Proton (150 MeV)	AAA, (Eclipse™, Varian) Electron MC Proton CS	CT <sub>cor</sub> (O-MAR) vs CT <sub>cor</sub> (kerMAR)	Depth/Range Absolute difference (Depth/Range) (Mean ± STD)	Refer to fig. 7 in Ref. [84] 0.7 ± 0.5 mm 1.3 ± 0.3 mm 1.8 ± 0.4 mm	kerMAR increased the particle range calculation accuracy
Liugang et al. <sup>81</sup>	MVCBCT & KVCT method	Phantom	SS	Photons (6 MV)	AAA, (Eclipse™, Varian)	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> & CT <sub>ref</sub> )	Maximum dose errors (%) Through hole Through phantom & metal	5.1% & 11.8% 3.3% & 17.2%	The proposed method improved the dose calculation accuracy
Ziemann et al. <sup>92</sup>	ALIR	Phantom	SS	Photons, 6 MV (VMAT)	PB and AAA (Eclipse™, Varian) [ionization chamber]	(CT <sub>art</sub> vs Measured) & (CT <sub>cor</sub> vs Measured)	Maximum dose errors (%)	8.4% & 2.7%	ALIR reduced the errors in dose calculation
Dong et al. <sup>72</sup>	Image-based MAR	Clinical (Hip); (n = 1) Clinical (H&N); (n = 1)	Simulated artifacts Dental fillings	Photons	N/A	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	%GP failure	23.25% & 0.02%	The proposed method increased the dose calculation accuracy
Odlozlikova et al. <sup>93</sup>	MDT	Clinical (Breast and lymph nodes); (n = 1), (Esophagus); (n = 2), (H& N); (n = 2), (breast); (n = 2), (Lung); (n = 2)	CIEDs	Photon (3D-CRT, IMRT, VMAT, and SBRT)	Eclipse™, Varian	CT <sub>art</sub> vs CT <sub>cor</sub>	The greatest dose error to ECIDs	>3% of total calculated dose	The application of MDT reduced the dose to the cardiac devices as OARs.
Aziz et al. <sup>73</sup>	Sinogram-based MAR	Phantom	Dental amalgam	Single-electron beam (BEAMnrc), 9MeV	MC, DOSXYZnrc source code	CT <sub>art</sub> vs CT <sub>cor</sub>	Dose error uncertainties at d80	56% & 2%	This MAR in combination with MC dose calculation reduced the dose calculation uncertainties
Jin et al. <sup>77</sup>	MAR with hardware adaptation	Clinical (Hip); (n = 1)	Simulated bilateral Ti implants	Proton	N/A	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> & CT <sub>ref</sub> )	Mean ΔWET	7mm & <1mm	The proposed MAR reduced the proton range errors

Ti Titanium, SS Stainless Steel, CRT Conformal Radiation Therapy, IMRT Intensity-Modulated Radiation Therapy, VMAT Volumetric-Modulated Arch Therapy, AAA Analytical anisotropic algorithm, PB Pencil Beam, MC Monte-Carlo, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, vs versus, ECIDs Electronic cardiac implantable devices, ΔWET Error in water equivalent thickness, %GP Gamma passing rate, N/A Not available.

The MRI-based MAR algorithm<sup>79</sup> was evaluated to assess its ability to improve the proton range error ( $\Delta$ WET) for proton therapy applications. On clinical CT scans of a brain and H&N, the application of this proposed method improved the absolute  $\Delta$ WET from 2.4 cm in brain and 1.7 cm in H&N to less than 2 mm for both cases. In a study by Nielsen et al.,<sup>84</sup> the performance of kerMAR was studied and compared with O-MAR for dosimetric impact in photon, electron, and proton therapy. In the phantom evaluation, a depth/range calculation on  $CT_{cor}$  for the proton, photons, and electrons did not show any significant difference ( $P > 0.05$ ) between the O-MAR and kerMAR. However, an evaluation of the absolute particle range difference on clinical  $CT_{cor}$  showed that kerMAR significantly ( $P < 0.001$ ) improved for electron beams nearly by 1 mm and proton beams by 2 mm in comparison with O-MAR. However, the photon beam evaluation showed an insignificant difference in depth/range calculations on  $CT_{cor}$ .

In a MVCBCT- and kVCT-based MAR study,<sup>81</sup> a custom-made phantom with a SS metal rod containing an air-filled hole in its center was CT scanned. The MAR method was applied to generate a  $CT_{cor}$  from  $CT_{art}$ . Subsequently, dose calculations were performed on  $CT_{art}$ ,  $CT_{cor}$ , and  $CT_{ref}$  using a single irradiation field in two instances, one crossing the hole and the other one crossing only the phantom and the metal. The calculated maximum percentage of dose error through the hole was 5.1% on  $CT_{cor}$  and 11.8% on  $CT_{art}$  in comparison with  $CT_{ref}$ . The dose error through the phantom and the metal was 3.3% on  $CT_{cor}$  and 17.2% on  $CT_{art}$ . So, this seems to imply that the calculated dose on the  $CT_{cor}$  after application of this MAR method was improved and better approximated the dose distribution on  $CT_{ref}$ .

ALIR was studied by Ziemann et al.<sup>92</sup> To simulate a pelvis with bilateral hip prostheses, a polymethylmethacrylate phantom with two SS rods was CT scanned. Then, ALIR, linear interpolation (LI), and density correction MAR methods were applied to create corresponding  $CT_{cor}$ . Next, a PTV was chosen at the center of the phantom between the metal implants and dose calculations were performed on  $CT_{cor}$ . Dose errors were computed between the calculated doses and respective measured doses using ion chambers. These dose errors were 2.7% for ALIR, 3.2% for LI, and 4.1% for the density correction. The  $CT_{art}$  showed a higher dose error of 8.4%. However, calculated dose to the simulated rectum as an OAR was higher for the ALIR method than for the LI method. The study concluded that ALIR can be applied for successful RT planning.

The image-based MAR<sup>72</sup> was assessed for the dosimetric impact on clinical H&N CT scans with dental fillings and on a CT scan of a pelvis with simulated hip implants. The calculation of gamma failure rates for a 3%/3 mm gamma criterion on  $CT_{art}$  and  $CT_{cor}$  compared to  $CT_{ref}$  resulted in 23.25% and 0.02%, respectively. In clinical studies, Odlozilikova et al. evaluated the MDT algorithm for dose error to cardiac implantable electronic devices (CIEDs) as an OAR while planning to irradiate the adjacent target tissues.<sup>93</sup> The total planned dose to the PTV was used to calculate the dose error to the CIEDs on  $CT_{art}$  and  $CT_{cor}$ . The largest dose error for the CIEDs was on the  $CT_{art}$ , and it was more than 3% of the total planned dose. Hence,

the application of MDT significantly improved RT planning and reduced the dose error to the CIEDs.

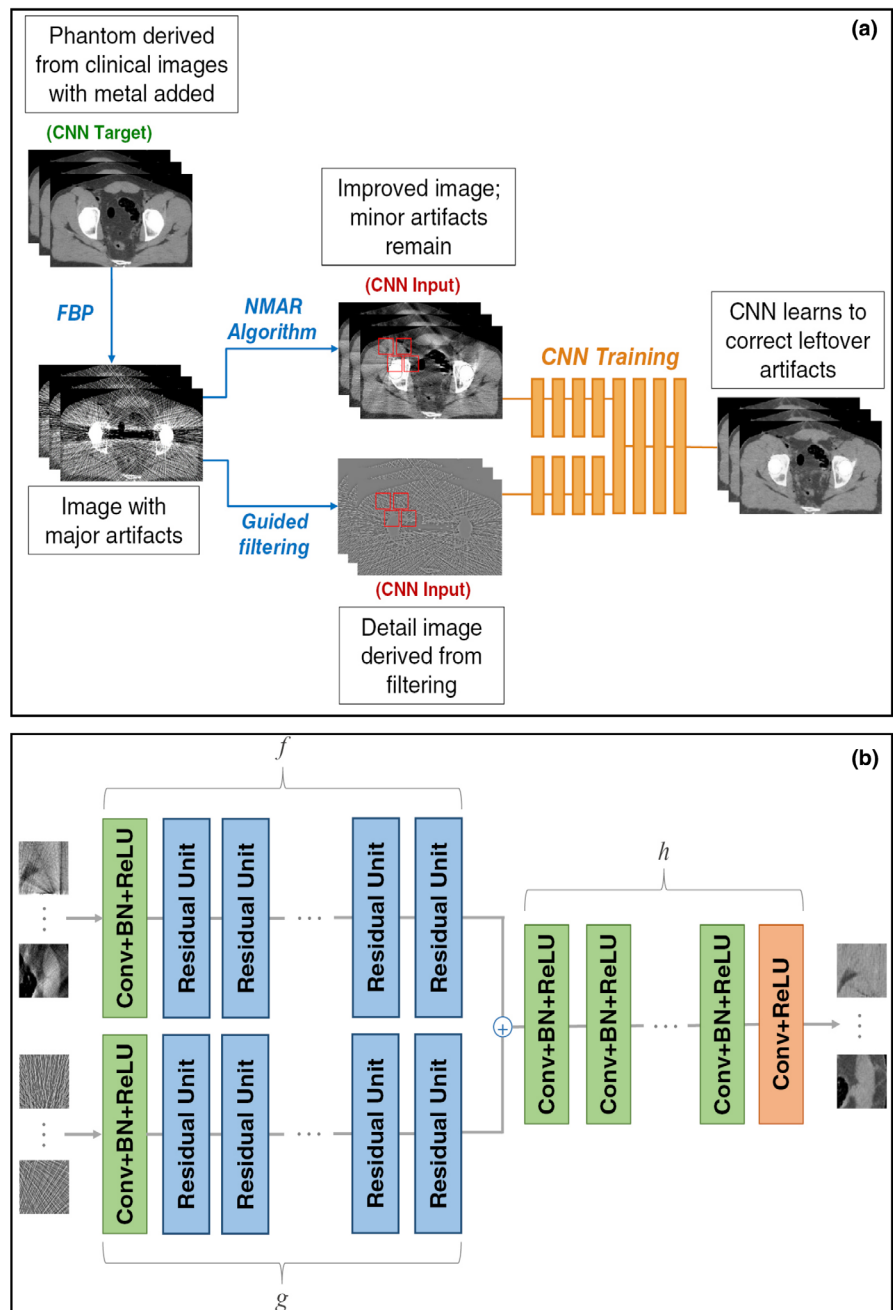
Aziz et al. evaluated dose calculation errors induced by a dental amalgam in a custom-made phantom experiment.<sup>73</sup> An in-house developed LI was applied to correct the metal artifacts, and a simulated electron beam was used to evaluate the dosimetric errors between the dose calculation plans on  $CT_{art}$  and  $CT_{cor}$ . Uncertainties in the calculated dose error reduced from 46% on  $CT_{art}$  to 2% on  $CT_{cor}$  at  $d_{80}$ .<sup>94</sup> For this reason, it was suggested to apply the proposed method to improve the dose calculation accuracy in electron beam therapy. In another study, clinical CT scans of a hip with simulated Ti implants were used for dose calculation on  $CT_{art}$ ,  $CT_{cor}$  after application of MAR method by Jin et al.,<sup>77</sup> and  $CT_{ref}$ . The proposed MAR method reduced the proton mean  $\Delta$ WET from 7 mm on  $CT_{art}$  to less than 1 mm on  $CT_{cor}$  in comparison with  $CT_{ref}$ .

### 3.B.2 | Deep learning-based MAR algorithms

Few research studies using deep learning-based methods to reduce metal artifacts on CT scans for RT applications were found in literature. These deep learning MAR approaches are either based on 2D residual learning-based convolutional neural networks (CNNs)<sup>95,96</sup> or 2D cycle generative adversarial networks (GANs).<sup>97</sup>

A dual-stream deep network was proposed by Gjestebj et al. which utilizes the residual learning-based CNNs (see Fig. 5).<sup>96</sup> This network requires paired data,  $CT_{art}$  and  $CT_{ref}$ , during the artifact reduction. First, a custom implementation of normalized MAR (NMAR)<sup>98</sup> and guided filtering<sup>99</sup> are applied to the artifact corrupted CT slice to reduce the metal artifacts and preserve the edge information of a CT slice, respectively. Subsequently, patches [indicated by 'red' squares in Fig. 5(a)] are extracted from these CT scans and are input into two identical parallel streams  $f$  and  $g$  [Fig. 5 (b)]. Each of these parallel streams starts with an initial parameter layer and is followed by 20 residual units. This network structure is referred to as DestreakNet.<sup>100</sup> Each residual unit consists of two convolutional layers (Conv) followed by batch normalization (BN) and a rectified linear unit (ReLU) [Fig. 5(b)]. The residual units help the network to learn the residual errors and generate respective feature maps. Finally, the feature maps from these streams are summed and passed together through  $h$  [Fig. 5(b)] to produce patches with metal artifact correction. The stream  $h$  contains another eight parameter layers and a final layer without the BN. During training, DestreakNet is optimized using a mean square error (MSE) loss function and perceptual loss function.<sup>101</sup> The MSE loss function is used to calculate the pixel-by-pixel error between the output and the target, while the perceptual loss function mitigates oversmoothing which can result from the MSE loss function.

A similar study, also using a residual learning-based artifact reduction CNN (RL-ARCNN) and paired data for artifact reduction, was proposed for brachytherapy applications by Huang et al.<sup>95</sup> This network architecture consists of a CNN, and ReLU is applied to increase the nonlinearity of the input layer. In subsequent convolutional layers, BN is applied between the convolution and ReLU to



**FIG. 5.** Overview of the deep learning approach proposed by Gjestebj et al.<sup>96</sup> (a) The proposed dual-stream deep network for metal artifacts reduction. (b) Details of the CNN training of (a). (Reproduced from Gjestebj et al.<sup>96</sup>. © Institute of Physics and Engineering in Medicine. Reproduced by permission of IOP Publishing. All rights reserved).

achieve faster training. This network was trained for MAR through the residual learning method as proposed by He et al.<sup>102</sup>

The main limitation of the above both networks, dual-stream deep network and RL-ARCNN, is that they both need paired data for training. To overcome this challenge, Koike et al. developed a deep learning-based MAR (DL-MAR)<sup>97</sup> using CycleGANs in which U-Net and PatchGAN architecture of the pix2pix model<sup>103</sup> are used as generator and discriminator, respectively. The generator synthesizes  $CT_{cor}$  from  $CT_{art}$  and vice versa, while the discriminator aims to distinguish the synthetic  $CT_{cor}$  scans as fake. The authors made the following modifications from the paper published by Isola et al.<sup>103</sup>: an instance normalization layer<sup>104</sup> replaced the BN layer in the generator, the kernel size of convolutional layer was set to 4X4 with a

stride of 2 instead of the maximum pooling filter, the image size was expanded by using the upsampling filters and one filter convolution layer was applied followed by a tanh activation function as the last filter. For all the other layers, a leaky rectified linear unit (LeakyReLU) with a slope of 0.2 was used as an activation function. In this proposed method, the  $CT_{art}$  is translated into  $CT_{cor}$  using adversarial loss, cycle consistency loss,<sup>105</sup> and identity map loss which were introduced to regularize the generator during the conversion.

**Evaluation of HU values restoration and image quality**

The proposed dual-stream deep network, RL-ARCNN, and DL-MAR were evaluated for their ability to restore HU values and image quality. The findings of these studies are summarized in Table 7.

**TABLE 7** Summary of the studies which investigated the restoration of HU values and image quality of deep learning-based MAR algorithms for RT application.

Author	MAR	Images; No. of CT slices /samples (n) [for training & for MAR]	Metals	CT scans	ROI	Measurements	Results	Comments
Gjesteby et al. <sup>96</sup>	Dual-stream CNN with residual learning	Phantom derived clinical (Hip & Spine); [42 & 08]	Simulated Ti	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> ) CT <sub>ref</sub> , CT <sub>art</sub> & CT <sub>cor</sub>	Whole image Hip Spine Near to metal implants on Hip Spine	SSIM PSNR in dB Mean HU	0.2382 & 0.8262 0.6930 & 0.8723 9.1830 & 22.1685 15.5450 & 21.8480 80.56, 2.25 & 94.26 94.27, 121.73 & 89.18	The proposed method is highly effective to reduce metal artifacts
Huang et al. <sup>95</sup>	RL-ARCNN	Clinical (Prostate); [550 & 550]	Simulated & real artifacts	(CT <sub>cor</sub> vs CT <sub>ref</sub> ) & (CT <sub>art</sub> vs CT <sub>ref</sub> )	Whole image	PSNR in dB	38.09 & 25.38	RL-CNN eliminates the metal artifacts better than conventional image processing-based MAR methods
Koike et al. <sup>97</sup>	Cycle GAN	Clinical (H&N); [(n = 92) & (n = 15)]	Dental fillings	(CT <sub>art</sub> vs CT <sub>ref</sub> ) & (CT <sub>cor</sub> vs CT <sub>ref</sub> )	Oral cavity	Artifact index [mean ± STD]	13.2 ± 4.3 & 267.3 ± 113.7	The proposed method reduced the metal artifacts from H&N CT scans

H&N Head & Neck, Ti Titanium, CT<sub>art</sub> CT scans with artifacts, CT<sub>cor</sub> Corrected CT scans, CT<sub>ref</sub> Reference CT scans, vs versus, SSIM Structural similarity, PSNR Peak Signal-to-Noise Ratio.

The dual-stream deep network with residual learning<sup>96</sup> was evaluated on synthesized data. Two kinds of voxelized phantoms were created from volumetric data of the pelvis and spinal regions from the visible human project.<sup>106</sup> CT scans of these voxelized phantoms were simulated by an industrial CT simulation software<sup>107</sup> with and without a Ti implants to mimic hip prostheses and spinal implants. Then, the proposed MAR algorithm was trained to generate CT<sub>cor</sub> from these CT scans. The authors measured the mean HU near the metal implant on CT<sub>ref</sub>, CT<sub>art</sub>, and CT<sub>cor</sub> and the results were as follows: 80.56, 2.25, and 94.26 HU in the hip case and 94.27, 121.73, and 89.18 HU in the spine case, respectively. Also, calculations of SSIM and PSNR were performed for CT<sub>art</sub> and CT<sub>cor</sub> in comparison with CT<sub>ref</sub>. For the SSIM calculation, a value of 1 indicates the highest similarity between the compared CT scans, while 0 indicates the lowest similarity. The calculated SSIM were 0.2382 for CT<sub>art</sub> and 0.8262 for CT<sub>cor</sub> for the hip case, and 0.6930 for CT<sub>art</sub> and 0.8723 for CT<sub>cor</sub> for the pelvis case. In addition, the calculated PSNR were 9.1830 dB for CT<sub>art</sub> and 22.1685 dB for CT<sub>cor</sub> for the hip case, and 15.5450 dB for CT<sub>art</sub> and 21.8480 dB for CT<sub>cor</sub> for the pelvis case. These results indicate that the proposed dual-stream deep network with residual learning can be used to reduce the metal artifacts for RT applications. However, this proposed MAR algorithm heavily depends on the performance of NMAR, used for data preprocessing.

For the brachytherapy application, RL-ARCNN<sup>95</sup> was evaluated. During the training, paired clinical cervical CT scans which include CT<sub>ref</sub> and simulated artifacts on CT<sub>ref</sub>, and CT scans with residual artifacts were used. The CT scans with residual artifacts were obtained from the difference between CT<sub>ref</sub> and CT<sub>ref</sub> with simulated

artifacts. During the MAR evaluation, clinical cervical CT scans with implanted seeds were used (CT<sub>art</sub>). The calculated PSNR was 38.09 dB for CT<sub>cor</sub> after application of RL-ARCNN and 25.38 dB for CT<sub>art</sub>. So, the proposed method seems to reduce the metal artifacts.

DL-MAR was studied on CT scans with dental implants by Koike et al.<sup>97</sup> The Adam optimizer was used to train the network using unpaired clinical H&N CT scans with (CT<sub>art</sub>) and without (CT<sub>ref</sub>) the dental implant. The calculation of the artifact index<sup>22</sup> (mean ± STD) between CT<sub>cor</sub> and CT<sub>art</sub> resulted in significant differences ( $P < 0.001$ ), and it was 13.2 ± 4.3 on CT<sub>cor</sub> and 267.3 ± 113.7 on CT<sub>art</sub>. Therefore, this proposed MAR algorithm can be used to reduce the metal artifacts for RT applications. However, the authors stated that the proposed MAR algorithm was trained and modified for the H&N CT scans and can only be applied on these scans.

#### Dosimetric impact of deep learning-based MAR algorithms

During the evaluation of DL-MAR<sup>97</sup> for RT, clinical H&N CT scans with dental implants were used. A density correction method using water density (1.0 g cm<sup>-3</sup>) was used for the comparison. IMRT planning with 6MV photons was performed using seven fields. To calculate the dose for the oral cavity on CT<sub>art</sub> and CT<sub>cor</sub>, the AAA algorithm was used. DAH was used to evaluate the calculated doses on CT<sub>cor</sub> after application of DL-MAR and CT<sub>art</sub> in comparison with the CT<sub>cor</sub> after application of density correction. The maximum dose differences were -2.4% on CT<sub>cor</sub> after application of DL-MAR and -7.2% on CT<sub>art</sub>. Moreover, very small dosimetric differences were found between the calculated doses on CT<sub>cor</sub> after application of DL-MAR and CT<sub>cor</sub> after application of density correction. However,

**TABLE 8** Strengths and weaknesses of proposed and/or investigated MAR methods for RT applications.

MAR methods		Strength	Weakness
Commercial MAR methods	O-MAR, iMAR, Smart MAR & SEMAR <sup>35,41-57,59,62,63</sup>	Standard methods and are used routinely in the clinic. Applicable on a wide range of clinical cases for RT applications.	Incomplete removal of artifacts with HU errors. Induce new artifacts. Scanner specific.
	Density correction methods <sup>38,67,68,71</sup>	Available in TPS.	Manual methods. Operator needs specialized knowledge and experience for artifact identification and density overrides.
Research-based traditional MAR algorithms	Image based MAR <sup>72</sup>	Do not use sinogram data	Semi-automatic
	MDT <sup>93</sup>	Reduces the dose to CIEDs during various RT procedures.	May requires long processing time
	Sinogram-based MAR <sup>73</sup>	Reduces the dose errors from amalgam in H&N RT.	Applicable for minor streaks artifacts
	STR, 3D median filter and Sinogram-based correction <sup>74</sup>	STR removes the spot artifacts. 3D median filter and sinogram-based corrections reduce the dark and white streaks.	STR and 3D median filter use predefined HU values and are applicable only for brachytherapy. Sinogram-based corrections induce new artifacts.
	Additional tilted CT scan-based MAR <sup>76</sup>	Reduces the HU errors better than LI and NMAR.	Additional radiation burden.
	ALIR <sup>92</sup>	Improves the dose calculation accuracy more than density correction method and LI do.	Increases the calculated doses for OAR.
	MAR with hardware adaptation <sup>77</sup>	Addresses the photon starvation during the artifact reduction.	Modified CT image acquisition in comparison with the standard.
	MRI-based CT MAR <sup>79</sup>	Does not require sinogram and threshold-based tissue classification.	Requires aligned MRI & CT scan. Semi-automatic.
	kerMAR <sup>84</sup>	Reduces the metal artifacts better than O-MAR in H&N cases.	Requires aligned MRI & CT scan. Applicable only for H&N case.
MVCBCT & kVCT method <sup>81</sup>	Performs better than LI and NMAR for artifact reduction.	Requires MVCBCT scan.	
Hybrid sinogram-based MAR <sup>82</sup>	Performs better than LI for artifact reduction.	Requires MVCT scan.	
Deep learning-based MAR algorithms	Dual-stream CNN with residual learning <sup>96</sup>	Reduces the remaining metal artifacts on CT scans after NMAR application.	Requires paired data and depends on the performance of NMAR
	RL-ARCNN <sup>95</sup>	Does not require sinogram data.	Needs paired data.
	DL-MAR <sup>97</sup>	Does not require paired data. Comparable performance to density correction for accuracy in dose calculation.	Applicable only in H&N cases.

the plan dose distribution on  $CT_{cor}$  after application of density correction may not reflect the actual dose distribution.

## 4 | DISCUSSION

Table 8 lists the strength and weakness of the proposed or evaluated MAR methods for RT applications. Commercial MAR algorithms are used in clinical environments to improve the treatment delivery in wide range of cases. These algorithms are self-optimized (automatic) and do not require skills from an operator. On the other hand,

in density correction methods, metal artifacts are identified by an expert and corrected by appropriate density overrides. Among the investigated research-based traditional MAR algorithms, the application of kerMAR reduced the metal artifacts more efficiently for H&N cases than O-MAR did. Even though ALIR improved the dose calculation accuracy more than the density correction method, it failed to reduce the planned radiation dose to certain OARs. MRI-based CT MAR<sup>79</sup> requires manual selection of CT scans for artifact correction. Most of the proposed MAR algorithms based on traditional image processing such as MVCT-based MAR, additional tilted CT scan-based MAR, and ALIR were performed better than LI or/and NMAR

for artifact correction. Among the deep learning-based MAR algorithms, DL-MAR does not require paired data for artifact correction, and it provided similar results for dose calculation when compared with the water density override.

The application of MAR methods on CT scans for RT applications was summarized in this review. Commercial MAR methods are utilized in clinical environments to correct the metal artifacts and improve the dosimetric accuracy in RT. However, residual artifacts and/or creation of new artifacts which may negatively impact in RT planning are often identified. Also, planned dose distributions on CT scans after application of TPS-based density correction methods show large-dose discrepancies compared to delivered dose. Thus, the performance of commercial MAR methods is not always completely satisfactory in RT application.

Among the research-based MAR methods, some of the MAR methods which are based on traditional image processing require an additional CT or aligned MRI scan which may result in extra radiation burden for the patient, or which may not be accessible. Only few studies compared the efficiency of the proposed MAR methods with the commercial MAR methods for RT applications. In addition, the ability of the proposed deep learning-based MAR algorithms to correct artifacts depends on the amount of CT scans from a specific anatomy which are used to train them. Thus, they are typically optimized to reduce the metal artifacts for a specific anatomy. Furthermore, their dosimetric impacts for RT applications were not extensively evaluated.

The commercial MAR methods and research-based MAR methods still have a limitation in metal artifact correction and/or dose improvements for RT applications. Their ability of the metal artifact corrections mainly depends on the anatomical region of CT scans which are corrupted by the metal artifacts from a specific implanted metal. The patterns and severity of the metal artifacts from each of implants are unique and different. The traditional MAR methods have difficulties in recognizing the patterns of the metal artifacts, but deep learning approaches can efficiently handle this complicated challenge. Also, newly induced artifacts were identified on CT scans after application of MAR algorithms which utilize the sinogram for the metal artifacts reduction. The deep learning-based MAR algorithms often work on reconstructed CT scans and, thus, the artifacts-corrected CT scans are free from these induced artifacts. Moreover, recent developments of MAR algorithms which utilize deep learning, for example, Cycle GANs do not require paired clinical CT scans for artifact reduction. Therefore, developing a MAR method while targeting a specific pattern of metal artifacts and a specific anatomical structure using a deep learning approach will be a promising solution. This method should be explored further and then evaluated for the RT applications.

## CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

## AUTHORS CONTRIBUTIONS

Sathyathas Puvanasunthararajah was involved in conceptualization, data extraction, and drafting of the manuscript. Davide Fontanarosa and Marie-Luise Wille were involved in conceptualization, reviewing, and supervision. Saskia M. Camps was involved in conceptualization, and checked the accuracy of data extraction, reviewing, and supervision.

## REFERENCES

1. De Man B, Nuyts J, Dupont P, Marchal G, Suetens P. Metal streak artifacts in X-ray computed tomography: a simulation study. *IEEE Trans Nucl Sci.* 1999;46:691–696.
2. Yadava GK, Pal D, Hsieh J. Reduction of metal artifacts: beam hardening and photon starvation effects, *Med. Imaging 2014 Phys. Med Imaging.* 2014;9033:90332V.
3. Kataoka ML, Hochman MG, Rodriguez EK, Lin PJP, Kubo S, Raptopoulos VD. A review of factors that affect artifact from metallic hardware on multi-row detector computed tomography. *Curr Probl Diagn Radiol.* 2010;39:125–136.
4. Stradiotti P, Curti A, Castellazzi G, Zerbi A. Metal-related artifacts in instrumented spine. Techniques for reducing artifacts in CT and MRI: state of the art. *Eur Spine J.* 2009;18:102–108.
5. Wellenberg RHH, Hakvoort ET, Slump CH, Boomsma MF, Maas M, Streekstra GJ. Metal artifact reduction techniques in musculoskeletal CT-imaging. *Eur J Radiol.* 2018;107:60–69.
6. Boas FE, Fleischmann D. CT artifacts: causes and reduction techniques. *Imaging Med.* 2012;4:229–240.
7. Katsura M, Sato J, Akahane M, Kunimatsu A, Abe O. Current and novel techniques for metal artifact reduction at CT: practical guide for radiologists. *Radiographics.* 2018;38:450–461.
8. Park HS, Chung YE, Seo JK. Computed tomographic beam-hardening artefacts: Mathematical characterization and analysis. *Philos Trans R Soc A Math Phys Eng Sci.* 2015;373:1–11.
9. Kim Y, Tomé WA, Bal M, McNutt TR, Spies L. The impact of dental metal artifacts on head and neck IMRT dose distributions. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* 2006;79:198–202.
10. Giantsoudi D, De Man B, Verburg J, et al. Metal artifacts in computed tomography for radiation therapy planning: dosimetric effects and impact of metal artifact reduction. *Phys Med Biol.* 2017;62:R49–R80.
11. Meyer E, Raupach R, Lell M, Schmidt B, Kachelrieß M. Frequency split metal artifact reduction (FSMAR) in computed tomography. *Med Phys.* 2012;39:1904–1916.
12. Meyer E, Raupach R, Lell M, Schmidt B, Kachelrieß M. Normalized metal artifact reduction (NMAR) in computed tomography. *Med Phys.* 2010;37:5482–5493.
13. Jin P, Bouman CA, Sauer KD. A model-based image reconstruction algorithm with simultaneous beam hardening correction for x-ray CT. *IEEE Trans Comput Imaging.* 2015;1:200–216.
14. Boudabbous S, Arditi D, Paulin E, Syrogiannopoulou A, Becker C, Montet X. Model-Based Iterative Reconstruction (MBIR) for the reduction of metal artifacts on CT. *Am J Roentgenol.* 2015;205:380–385.
15. Shah P, Kendall F, Khozin S, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. *Npj Digit Med.* 2019;2:69.
16. Levina AB, Schlosser C, Grewal J, Coope R, Jones SJM, Yip S. Rise of the machines: advances in deep learning for cancer diagnosis. *Trend Cancer.* 2019;5:157–169.

17. Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine learning and deep learning in medical imaging: intelligent imaging. *J Med Imaging Radiat Sci.* 2019;50:477–487.
18. Wang G, Ye JC, De Man B. Deep learning for tomographic image reconstruction. *Nat Mach Intell.* 2020;2:737–748.
19. Gjestebly L, Yang Q, Xi Y, Claus BEH, Jin Y, De Man B, Wang G, Shan H. Deep learning methods for CT image-domain metal artifact reduction. 2017;31. <https://doi.org/10.1117/12.2274427>.
20. Yu H, Zhang Y, Chu Y. Reduction of metal artifacts in x-ray CT images using a convolutional neural network. 2017, 30. <https://doi.org/10.1117/12.2275592>.
21. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in X-ray computed tomography. *IEEE Trans Med Imaging.* 2018;37:1370–1381.
22. Dong Y, Shi AJ, Wu JL, et al. Metal artifact reduction using virtual monochromatic images for patients with pedicle screws implants on CT. *Eur Spine J.* 2016;25:1754–1763.
23. Hanna GG, Hounsell AR, O'Sullivan JM. Geometrical analysis of radiotherapy target volume delineation: a systematic review of reported comparison methods. *Clin Oncol.* 2010;22:515–525.
24. Verdun FR, Racine D, Ott JG, et al. Image quality in CT: from physical measurements to model observers. *Phys Medica.* 2015;31:823–843.
25. Boone JM, Brink JA, Edyvean S, et al. 11. noise assessment in CT. *J ICRU.* 2012;12:121–134.
26. Renieblas GP, Nogués AT, González AM, Gómez-Leon N, del Castillo EG. Structural similarity index family for image quality assessment in radiological images. *J Med Imaging.* 2017;4:035501.
27. Li H, Dong L, Zhang L, Yang JN, Gillin MT, Zhu XR. Toward a better understanding of the gamma index: investigation of parameters with a surface-based distance method. *Med Phys.* 2011;38:6730–6741.
28. A C, C D, J B. The use of gamma indices with medical imaging as quality assurance tool to validate the dose calculation algorithm in the modern practice of medical physics. *Nucl Med Biomed Imaging.* 2016;1:31–34.
29. Astrahan MA. The retina dose-area histogram: a metric for quantitatively comparing rival eye plaque treatment options. *J Contemp Brachytherapy.* 2013;5:23–32.
30. Sale CA, Yeoh EEK, Scutter S, Bezak E. 2D versus 3D radiation therapy for prostate carcinoma: a direct comparison of dose volume parameters. *Acta Oncol.* 2005;44:348–354.
31. Zhang R, Taddei PJ, Fitzek MM, Newhauser WD. Water equivalent thickness values of materials used in beams of protons, helium, carbon and iron ions. *Phys Med Biol.* 2010;55:2481–2493.
32. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ.* 2009;339:b2700.
33. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.
34. Metal Artifact Reduction for Orthopedic Implants (O-MAR), Philips Healthc. 2011;1–12.
35. Axente M, Paidi A, Von Eyben R, et al. Clinical evaluation of the iterative metal artifact reduction algorithm for CT simulation in radiotherapy. *Med Phys.* 2015;42:1170–1183.
36. GE Healthcare. Smart metal artifact reduction (MAR), Europe. 2013;2.
37. Gondim Teixeira PA, Meyer J-B, Baumann C, et al. Total hip prosthesis CT with single-energy projection-based metallic artifact reduction: impact on the visualization of specific periprosthetic soft tissue structures. *Skeletal Radiol.* 2014;43:1237–1246.
38. Maerz M, Koelbl O, Dobler B. Einfluss von metallischen Zahnimplantaten und Metallartefakten auf die Genauigkeit der Dosisberechnung. *Strahlentherapie Und Onkol.* 2015;191:234–241.
39. Papanikolaou N, Battista JJ, Boyer AL, Kappas C, Klein E, Mackie TR. AAPM report 85:Tissue Inhomogeneity Corrections for Megavoltage Photon Beams. Report of the AAPM radiation therapy committee task group 65. 2004. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:TISSUE+INHOMOGENEITY+CORRECTIONS+FOR+MEGAVOLTAGE+PHOTON+BEAMS#0>.
40. Kamalian S, Lev MH, Gupta R. Chapter 1 - Computed tomography imaging and angiography – principles. In: Masdeu JC, González RG eds. *Neuroimaging Part I.* Amsterdam, Netherlands: Elsevier; 2016:3–20.
41. Kwon H, Kim KS, Chun YM, et al. Evaluation of a commercial orthopaedic metal artefact reduction tool in radiation therapy of patients with head and neck cancer. *Br J Radiol.* 2015;88:13–22.
42. Bär E, Schwahofer A, Kuchenbecker S, Häring P. Improving radiotherapy planning in patients with metallic implants using the iterative metal artifact reduction (iMAR) algorithm. *Biomed Phys Eng Express.* 2015;1:025206.
43. Maerz M, Mittermair P, Krauss A, Koelbl O, Dobler B. Iterative Metallartefakt Reduktion verbessert die Genauigkeit der Dosisberechnung: phantomstudie mit Zahnimplantaten. *Strahlentherapie Und Onkol.* 2016;192:403–413.
44. Guilfoile C, Rampant P, House M. The impact of smart metal artefact reduction algorithm for use in radiotherapy treatment planning. *Australas Phys Eng Sci Med.* 2017;40:385–394.
45. Huang VW, Kohli K. Evaluation of new commercially available metal artifact reduction (MAR) algorithm on both image quality and relative dosimetry for patients with hip prosthesis or dental fillings. *Int J Med Phys, Clin Eng Radiat Oncol.* 2017;06:124–138.
46. Inal A, Sarpün IH. Dosimetric evaluation of phantoms including metal objects with high atomic number for use in intensity modulated radiation therapy. *Radiat Environ Biophys.* 2020;59:503–510.
47. Murazaki H, Fukunaga J, Hirose T-A, et al. Dosimetric assessment of a single-energy metal artifact reduction algorithm for computed tomography images in radiation therapy. *Radiol Phys Technol.* 2019;12:268–276.
48. Miki K, Mori S, Hasegawa A, Naganawa K, Koto M. Single-energy metal artefact reduction with CT for carbonion radiation therapy treatment planning. *Br J Radiol.* 2016;89:20150988.
49. Sillanpää J, Lovelock M, Mueller B. The effects of the orthopedic metal artifact reduction (O-MAR) algorithm on contouring and dosimetry of head and neck radiotherapy patients. *Med Dosim.* 2020;45:92–96.
50. Andersson KM, Dahlgren CV, Reizenstein J, Cao Y, Ahnesjö A, Thunberg P. Evaluation of two commercial CT metal artifact reduction algorithms for use in proton radiotherapy treatment planning in the head and neck area. *Med Phys.* 2018;45:4329–4344.
51. Kohan AA, Rubbert C, Vercher-Conejero JL, et al. The impact of orthopedic metal artifact reduction software on interreader variability when delineating areas of interest in the head and neck. *Pract Radiat Oncol.* 2015;5:e309–e315.
52. Hansen CR, Christiansen RL, Lorenzen EL, et al. Contouring and dose calculation in head and neck cancer radiotherapy after reduction of metal artifacts in CT images. *Acta Oncol (Madr).* 2017;56:874–878.
53. Hagen M, Kretschmer M, Würschmidt F, et al. Clinical relevance of metal artifact reduction in computed tomography (iMAR) in the pelvic and head and neck region: multi-institutional contouring study of gross tumour volumes and organs at risk on clinical cases. *J Med Imaging Radiat Oncol.* 2019;63:842–851.
54. Kovacs DG, Rechner LA, Appelt AL, et al. Metal artefact reduction for accurate tumour delineation in radiotherapy. *Radiother Oncol.* 2018;126:479–486.



55. Shiraiishi Y, Yamada Y, Tanaka T, et al. Single-energy metal artifact reduction in postimplant computed tomography for I-125 prostate brachytherapy: impact on seed identification. *Brachytherapy*. 2016;15:768–773.
56. Huang JY, Followill DS, Howell RM, et al. Approaches to reducing photon dose calculation errors near metal implants. *Med Phys*. 2016;43:5117–5130.
57. Shen ZL, Xia P, Klahr P, Djemil T. Dosimetric impact of orthopedic metal artifact reduction (O-MAR) on spine SBRT patients. *J Appl Clin Med Phys*. 2015;16:106–116.
58. Abel S, Lee S, Ludmir EB, Verma V. Principles and applications of stereotactic radiosurgery and stereotactic body radiation therapy. *Hematol Oncol Clin North Am*. 2019;33:977–987.
59. Akdeniz Y, Yegingil I, Yegingil Z. Effects of metal implants and a metal artifact reduction tool on calculation accuracy of AAA and Acuros XB algorithms in small fields. *Med Phys*. 2019;46:5326–5335.
60. Sievinen J, Ulmer W, Kaissl W. AAA photon dose calculation model in Eclipse, Palo Alto Varian Med. Syst. Varian doc. 2005;1–23. [http://www.rtsalon.cn/upload/RTsalon\\_p\\_3218\\_2.pdf](http://www.rtsalon.cn/upload/RTsalon_p_3218_2.pdf).
61. Elsherbini N, Ahmed W, Elwakeel S. Verification of performance of Acuros XB Algorithm (AXB) Implemented on Eclipse Planning System. *Res Oncol*. 2018;1–5.
62. Jia Y, Zhao L, Cheng C, McDonald M, Das I. SU-E-T-450: dose perturbation of spinal metal implants in proton beam therapy. *Med Phys*. 2013;40:309.
63. Righetto R, Clemens LP, Lorentini S, et al. Accurate proton treatment planning for pencil beam crossing titanium fixation implants. *Phys Medica*. 2020;70:28–38.
64. Taylor A, Powell MEB. Intensity-modulated radiotherapy - What is it? *Cancer Imaging*. 2004;4:68–73.
65. Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. *Br J Radiol*. 2011;84:967–996.
66. Antolak JA, Rosen II. Planning target volumes for radiotherapy: how much margin is needed? *Int J Radiat Oncol Biol Phys*. 1999;44:1165–1170.
67. Acquah GF, Kyeremeh PO, Hasford F, Boadu M, Sosu EK, Inkoom S. Evaluation of metallic implant artifact on photon beam calculation algorithms using a CIRS thorax phantom. *J Radiat Res Appl Sci*. 2018;11:347–349.
68. Parenica HM, Mavroidis P, Jones W, Swanson G, Papanikolaou N, Stathakis S. VMAT optimization and dose calculation in the presence of metallic hip prostheses. *Technol Cancer Res Treat*. 2019;18:1–10.
69. McNutt T. Dose calculations collapsed cone convolution superposition and delta pixel beam, Pinnacle White Pap. 2002;1–10.
70. Jabbari K. Review of fast monte carlo codes for dose calculation in radiation therapy treatment planning. *J Med Signals Sens*. 2011;1:73–86.
71. Parenica HM, Ford JR, Mavroidis P, Li Y, Papanikolaou N, Stathakis S. Treatment planning dose accuracy improvement in the presence of dental implants. *Med Dosim*. 2019;44:159–166.
72. Dong X, Yang X, Rosenfield J, Elder E, Dhabaan A. Image-based metal artifact reduction in X-ray computed tomography utilizing local anatomical similarity. 2019;1–12. <https://doi.org/10.1117/12.2255083>. Image-based.
73. Aziz MZA, Yusoff AL, Osman ND, Abdullah R, Rabaie NA, Salikin MS. Monte carlo dose calculation in dental amalgam phantom. *J Med Phys*. 2015;40:150–155.
74. Miksys N, Xu C, Beaulieu L, Thomson RM. Development of virtual patient models for permanent implant brachytherapy Monte Carlo dose calculations: interdependence of CT image artifact mitigation and tissue assignment. *Phys Med Biol*. 2015;60:6039–6062.
75. Boas FE, Fleischmann D. Evaluation of two iterative techniques for reducing metal artifacts in computed tomography. *Radiology*. 2011;259:894–902.
76. Kim C, Pua R, Lee C-H, et al. An additional tilted-scan-based CT metal-artifact-reduction method for radiation therapy planning. *J Appl Clin Med Phys*. 2019;20:237–249.
77. Jin Y, Giantsoudi D, Fu L, et al. Metal artifact reduction for radiation therapy: a simulation study. 2018;25. <https://doi.org/10.1117/12.2293840>.
78. Stille M, Kleine M, Hägele J, Barkhausen J, Buzug TM. Augmented likelihood image reconstruction. *IEEE Trans Med Imaging*. 2016;35:158–173.
79. Park PC, Schreiber E, Roper J, et al. MRI-based computed tomography metal artifact correction method for improving proton range calculation accuracy. *Int J Radiat Oncol Biol Phys*. 2015;91:849–856.
80. Nielsen JS, Edmund JM, Van Leemput K. Magnetic resonance-based computed tomography metal artifact reduction using Bayesian modelling. *Phys Med Biol*. 2019;64:245012.
81. Liugang G, Hongfei S, Xinye N, Mingming F, Zheng C, Tao L. Metal artifact reduction through MVCBCT and kVCT in radiotherapy. *Sci Rep*. 2016;6:1–11.
82. Jeon H, Park D, Youn H, et al. Generation of hybrid sinograms for the recovery of kV-CT images with metal artifacts for helical tomotherapy. *Med Phys*. 2015;42:4654–4667.
83. Kalender WA, Hebel R, Ebersberger J. Reduction of CT artifacts caused by metallic implants. *Radiology*. 1987;164:576–577.
84. Nielsen JS, Van Leemput K, Edmund JM. MR-based CT metal artifact reduction for head-and-neck photon, electron, and proton radiotherapy. *Med Phys*. 2019;46:4314–4323.
85. Wilks DS. Bayesian inference. *Int Geophys*. 2011;100:187–213.
86. Wang J, Wang S, Chen Y, Wu J, Coatrieux J-L, Luo L. Metal artifact reduction in CT using fusion based prior image. *Med Phys*. 2013;40:81903.
87. Yu DF, Fessler JA, Ficarò EP, Yu DF, Ficarò EP. Maximum-likelihood transmission image reconstruction for overlapping transmission beams. *IEEE Trans Med Imaging*. 2000;19:1094–1105.
88. Dayanand K. 乳鼠心肌提取 HHS public access. *Physiol Behav*. 2018;176:139–148.
89. Fu L, Lee T-C, Kim SM, et al. Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction. *IEEE Trans Med Imaging*. 2017;36:707–720.
90. Basran PS, Robertson A, Wells D. CT image artifacts from brachytherapy seed implants: a postprocessing 3D adaptive median filter. *Med Phys*. 2011;38:712–718.
91. Segars WP, Sturgeon G, Mendonca S, Grimes J, Tsui BMW. 4D XCAT phantom for multimodality imaging research. *Med Phys*. 2010;37:4902–4915.
92. Ziemann C, Stille M, Cremers F, Buzug TM, Rades D. Improvement of dose calculation in radiation therapy due to metal artifact correction using the augmented likelihood image reconstruction. *J Appl Clin Med Phys*. 2018;19:227–233.
93. Odložilíková A, Kurzyuková A, Sepsí M, Pospíšil D, Slampa P. Použití metody Metal Deletion Technique pro odstraňování artefaktů při plánování radioterapie u pacientů s implantovanými kardiostimulátory. *Klin Onkol*. 2018;31:434–438.
94. Hogstrom KR, Almond PR. Review of electron beam therapy physics. *Phys Med Biol*. 2006;51:R455–R489.
95. Huang X, Wang J, Tang F, Zhong T, Zhang Y. Metal artifact reduction on cervical CT images by deep residual learning 08 Information and Computing Sciences 0801 Artificial Intelligence and Image Processing. *Biomed Eng Online*. 2018;17:1–15.
96. Gjestebly L, Shan H, Yang Q, et al. A dual-stream deep convolutional network for reducing metal streak artifacts in CT images. *Phys Med Biol*. 2019;64:235003.
97. Koike Y, Anetai Y, Takegawa H, Ohira S, Nakamura S, Tanigawa N. Deep learning-based metal artifact reduction using cycle-consistent adversarial network for intensity-modulated head and neck radiation therapy treatment planning. *Phys Medica*. 2020;78:8–14.

98. Meyer E, Bergner F, Raupach R, Flohr T, Kachelrieß M. Normalized metal artifact reduction (NMAR) in computed tomography, in: 2009 IEEE Nucl. Sci. Symp. Conf. Rec. 2009:3251–3255.
99. Fu X, Huang J, Zeng D, et al. Removing Rain from Single Images via a Deep Detail Network, in: 2017 IEEE Conf. Comput Vis Pattern Recognit. 2017:1715–1723.
100. Gjestebj L, Shan H, Yang Q, et al. Deep neural network for CT metal artifact reduction with a perceptual loss function, In: 2018.
101. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. *IEEE Trans Comput Imaging*. 2017;3:47–57.
102. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016;770–778.
103. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks, Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017;5967–5976.
104. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization. 2016.
105. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks, Proc. IEEE Int. Conf. Comput. Vis. 2017;2242–2251.
106. Ackerman MJ. The visible human project. *Proc IEEE*. 1998;86:504–511.
107. De Man B, Basu S, Chandra N, et al. CatSim: a new computer assisted tomography simulation environment, in. Proc. SPIE. 2007.