Journal of Imaging Informatics in Medicine
https://doi.org/10.1007/s10278-026-02030-x

Deep Learning–Based Metal Artifact Reduction in Cardiac Computed Tomography: A Preliminary Study Enabling Radiomic Analysis in Patients with Implantable Defibrillators

Nicholas Benigni[1] · **Francesca Lo Iacono**[2] · **Lisa M. Verheul**[3] · **Marco Guglielmo**[3] · **Paul Volders**[3] · **Rachel ter Bekke**[3] · **Gianluca Pontone**[4,5] · **Rutger J. Hassink**[6] · **Valentina D. A. Corino**[1,2]

Received: 26 January 2026 / Revised: 7 May 2026 / Accepted: 23 May 2026

Abstract

Idiopathic ventricular fibrillation (IVF) affects 5–10% of out-of-hospital cardiac arrest survivors, requiring implantable cardioverter–defibrillators (ICDs) for management. However, metal artifacts from ICDs may compromise cardiac computed tomography (CCT) image quality, limiting diagnostic capabilities and preventing advanced analyses like radiomics. The aim of this study is to develop a deep learning (DL)-based solution to reduce metal-induced artifacts by effectively removing both the metallic implant (primary artifact) and the associated streak artifacts (secondary artifact) in CCT scans of IVF patients with ICDs and to evaluate its effectiveness through radiomic analysis for predicting a composite clinical endpoint. CCT scans from 41 IVF patients with ICDs and 13 unaffected family members without ICDs were included. A fully convolutional neural network with 20 layers was trained using a simulated dataset combining artifact-free images from controls with real artifact masks extracted from patient scans. The network was evaluated using structural similarity index (LV-SSIM), mean absolute error (LV-MAE), and mean squared error (LV-MSE), computed on the segmented region of interest (ROI), i.e., left ventricle wall. Further, 851 radiomic features were extracted from the ROI, and machine learning models were developed to predict a composite endpoint including nonsustained ventricular tachycardia, high premature ventricular contraction burden, and ventricular arrhythmias recurrences. The DL model achieved good performance with an LV-SSIM of 0.936 ± 0.045, LV-MAE of 15.59 ± 8.48 HU, and LV-MSE of 661.89 ± 1003.80 HU on the test set. Radiomic feature analysis demonstrated that 98% of stable features were preserved after artifact removal, confirming diagnostic integrity. The predictive model achieved an $F1$ score of 0.85 for the composite clinical endpoint, demonstrating effective results for metal artifact reduction in CCT of IVF patients with ICDs. A promising DL solution for metal artifact reduction in CCT imaging of patients with ICD has been proposed. The demonstrated preservation of radiomic features suggests that the approach maintains diagnostic integrity, potentially expanding the utility of CCT in patients with cardiac devices who previously might have been excluded from certain studies.

Keywords Cardiac computed tomography · Idiopathic ventricular fibrillation · Artifact reduction · Deep learning · Defibrillator

Introduction

Idiopathic ventricular fibrillation (IVF) is diagnosed in patients who survived a ventricular fibrillation episode without any identifiable structural or electrical cause after diagnostic investigations, including imaging, electrophysiological studies, and structural, clinical, pharmacological and genetic tests [1]. It is a more common cause of sudden death in young adults and accounts for 5 to 10% of survivors of out-of-hospital cardiac arrest [1, 2]. Due to the unpredictable nature and high recurrence rate of ventricular fibrillation episodes in these patients, implantable cardioverter–defibrillators (ICDs) are the recommended first-line therapy, providing life-saving shock delivery when dangerous arrhythmias recur [2].

Follow-up care is essential for both patients and their family members. For patients, regular monitoring is crucial as some may develop identifiable cardiac phenotypes over time, and advances in diagnostic techniques might reveal

Nicholas Benigni and Francesca Lo Iacono equally contributed to this work.

Extended author information available on the last page of the article

previously undetected abnormalities [1]. Family screening is important because some inherited arrhythmia syndromes initially diagnosed as idiopathic ventricular fibrillation (IVF) have genetic components that may affect relatives [1].

A significant challenge in the clinical management of IVF patients with ICDs is the image quality degradation during cardiac computed tomography (CCT). Metal components, such as defibrillators or pacemakers, generate severe artifacts through beam hardening, photon starvation, and scattering effects during CCT acquisition [3]. Metal objects absorb or block X-rays and produce incomplete data at the detector. When images are reconstructed using these compromised data, artifacts appear as characteristic bright or dark streaks that make it difficult to distinguish between different types of tissue [4]. These effects, highly dependent on the size, shape, and alloy of the implant [5], may compromise image quality and potentially lead to missed clinical findings [6] and reduced diagnostic value.

Metal artifact reduction (MAR) approaches have been developed to address these limitations in cardiac imaging [4, 5]. Early sinogram completion methods [7] identify metal-affected projections, remove corrupted data, and replace them using linear interpolation before reconstructing the image. While pioneering, these techniques often compromise edge information and introduce new artifacts, and they have not been applied extensively to CCT. Variants of sinogram completion and commercial software solutions, including Smart Metal Artifact Reduction (SEMAR) by Canon [3, 8, 9] and Orthopedic Metal Artifact Reduction (O-MAR) by Philips [10], are effective for orthopedic implants but show limited performance with cardiac implantable electronic devices (CIEDs) in CCT. Iterative reconstruction techniques have also been used to reduce blooming artifacts from coronary stents [11, 12], but these artifacts are minor compared to those caused by CIEDs. These limitations highlight the need for deep learning-based approaches capable of handling the complex and large artifacts encountered in cardiac imaging, motivating the present study. DL has recently emerged as a promising approach for reducing metal artifacts in CT images in orthopedic and dental areas, although cardiovascular applications remain almost underexplored, with one prior study using a three-CNN ensemble to address both metal and motion artifacts simultaneously in cardiac scans—the Pacemaker Artifact Reduction (DyPAR +) method [13] and a more recent study that proposed a generative adversarial network (GAN)-based approach, termed Cardiac Inpainting Net [14], which aims to fill image regions affected by metal implants and the associated artifacts.

The aim of this study is to reduce metal-induced artifacts by effectively removing both the metallic implant (primary artifact) and the associated streak artifacts (secondary artifact), thereby producing images suitable for radiomic analysis. In particular, the goal is to minimize nonphysiological textural patterns caused by these artifacts, which could otherwise bias feature extraction and compromise the reliability of radiomic analysis. To evaluate the effectiveness of this method a radiomic analysis is conducted on IVF patients to predict a composite endpoint including nonsustained ventricular tachycardia (NSVT), high premature ventricular contraction (PVC) burden, and ventricular arrhythmia recurrence.

Materials and Methods

Study Population

This study included 41 IVF patients (median age 42 [31–55], 61% male) with an ICD and 13 unaffected family members without ICDs (median age 27 [22–55], 54% male) of the IVF patients, defined as the control group, from the University Medical Center Utrecht and Maastricht University Medical Center + (MUMC +). During follow-up of patients with IVF, the following events were recorded, creating a composite endpoint: nonsustained ventricular tachycardia (NSVT), high premature ventricular contraction (PVC) burden, and arrhythmias that led to appropriate ICD therapy. This study complies with the Declaration of Helsinki and was approved by local institutional ethics review boards (NL67079.068.18/METC18-038).

Among the IVF patients, two did not have follow-up information, and three others had impaired image quality, with artifacts related to nonrespect of breath hold, premature ventricular contractions, blooming effect, motion artifacts related to heart rate variability, interference of cardiac veins, intramyocardial tract, and impaired signal/image noise ratio. Thus, the final number of analyzed IVF patients was 36. The IVF group included a total of 9152 images (254 ± 78 images per patient) while the control group consisted of 2883 images (222 ± 23 images per subject).

All scans were acquired as 512×512 matrices in 12-bit format. Left ventricular (LV) wall segmentation was performed using 3D Slicer software (version 5.6.2) with subsequent validation by an expert radiologist holding Level III certification from the European Association of Cardiovascular Imaging. Written informed consent was obtained from all participants prior to enrollment.

Dataset Generation

To train and evaluate a network for MAR, a dataset of paired artifact-free and CCT images with simulated artifacts was built. Artifact masks were extracted from the IVF group's scans, generating a total of 1469 candidate masks. Artifact-free images were selected from control group scans and artifact masks were added to these clean images, generating pairs of images with simulated artifacts and without artifacts that allowed for evaluation of the method's effectiveness (see an example in Fig. 1). Dataset generation steps are described in the following paragraphs.

Artifact-Free CCT Images

Images without artifacts were obtained from scans of the control group, focusing exclusively on slices containing the LV. A total of 628 images were obtained and employed to generate simulated images with artifacts.

Metal Mask

Metal masks were automatically segmented from IVF patients' CCT images by thresholding at 2800 Hounsfield Units (HU) as suggested in [15], obtaining 1469 candidate masks. To ensure proper positioning of simulated metal over cardiac tissue, masks' positioning was adjusted by applying a predefined set of rotations (range 10–100 degrees with 10-degree increments), clockwise and counterclockwise, and translations (range 10–100 pixels with 10-pixel increments) leftward and rightward. The optimal transformation was selected based on the overlap between mask pixels and cardiac tissue, identified as region with radiodensity values between 0 and 1000 HU. When different transformations yielded equivalent performance, rotation was prioritized, while larger displacement was selected when dealing with the same type of transformations. A random contouring [16] was performed to slightly modify mask boundaries, generating different metal shapes and noise profiles.

Metal Artifact Simulation

The simulation was based on the following steps as in Sakamoto et al. [17]: i) according to the intensity value, "water weight" and "bone weight" proportions were calculated for each voxel using a linear transition function between 100 HU (water) and 1500 HU (bone) [18]; ii) water-weighted and bone-weighted images were created through thresholding; iii) forward projection was performed on water-based and bone-based images, along with an additional image where the metal region was filled with the linear attenuation coefficient of titanium. Titanium was chosen as a primary noise

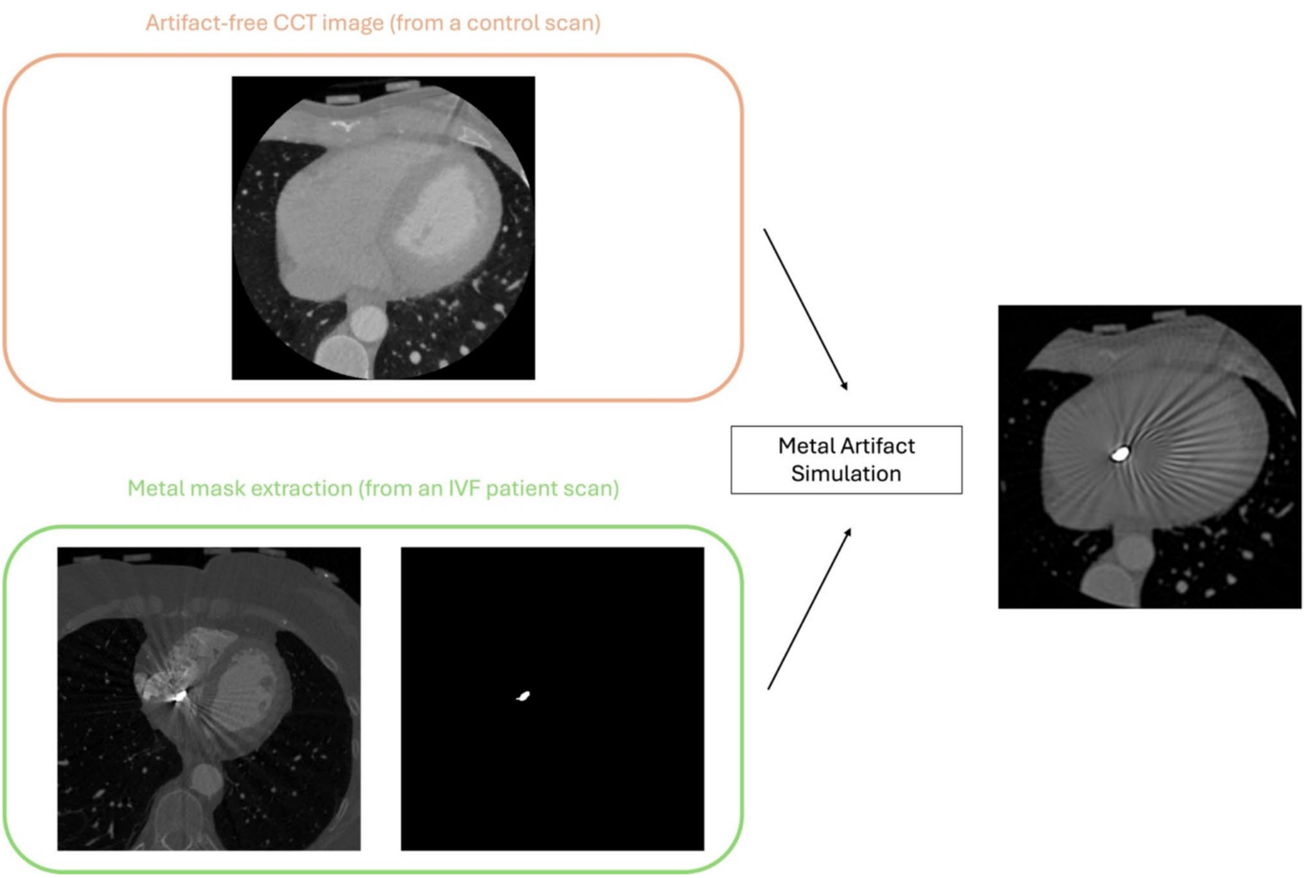

Fig. 1 Creation of simulated image with artifacts: artifact-free image (top panel), metal shape extracted from an image of a patient with idiopathic ventricular fibrillation (bottom panel)

source due to its prevalent use in medical applications. iv) polychromatic projections were computed using mass attenuation coefficients from the XCOM dataset [19], assuming a tungsten anode CT scanner with 120-kVp tube voltage. v) Poisson random noise was applied to simulate quantum noise. vi) After water beam-hardening correction, the final images were reconstructed using filtered back-projection with a Ram-Lak filter. vii) Images were converted to grayscale and normalized to a [0–4095] range for compatibility with standard CT imaging characteristics.

Simulated Image Dataset

To enhance dataset variability, five different metal masks were applied to each artifact-free slice, generating five different versions of simulated images with artifacts for each of them. The final simulated image dataset comprised 3140 CCT slices with artifacts, each paired with its original artifact-free image, used as a target for model training.

Metal Artifact Reduction DL Development

A fully convolutional DL model based on Kim et al. [20] was employed to reduce the metal artifact from the CCT images. The architecture consisted of 20 sequential convolutional layers with 3×3 kernels and ReLU activation functions. This kernel size was selected to capture fine-grained features while minimizing distortion risks from wider image regions. Each hidden layer contained 64 channels, while input and output layers used single channels to match image dimensions. A skip connection before the output layer allowed input image concatenation with final layer features, preserving structural integrity and facilitating gradient flow during training. Figure 2 reports a schematic view of the model.

Mean absolute error (MAE) was used as a loss function. Training was performed for 100 epochs with a batch size of 32 images using the Adam optimizer at a learning rate of 0.01. Data were shuffled after each epoch to ensure generalized learning. Training incorporated two callback functions: early stopping, which monitored structural similarity index (SSIM, see the next section for details) on the validation set with 20 epochs patience, and reduce learning rate on Plateau, which monitored SSIM with 7 epochs patience and a 0.4 reduction factor.

The preprocessing workflow focused on targeted regions rather than entire image slices. For each patient, a bounding box including the segmented ventricle was first identified, and the corresponding section of the slice was resized to a standardized 128×128 matrix. Images were thus normalized to a 0–1 value range. The dataset was divided using a patient-wise stratification strategy, with 11 patients assigned to the training set and 2 to the test set, resulting in an approximate 80–20% split. All images from a given patient were exclusively assigned to the same set. To mitigate potential

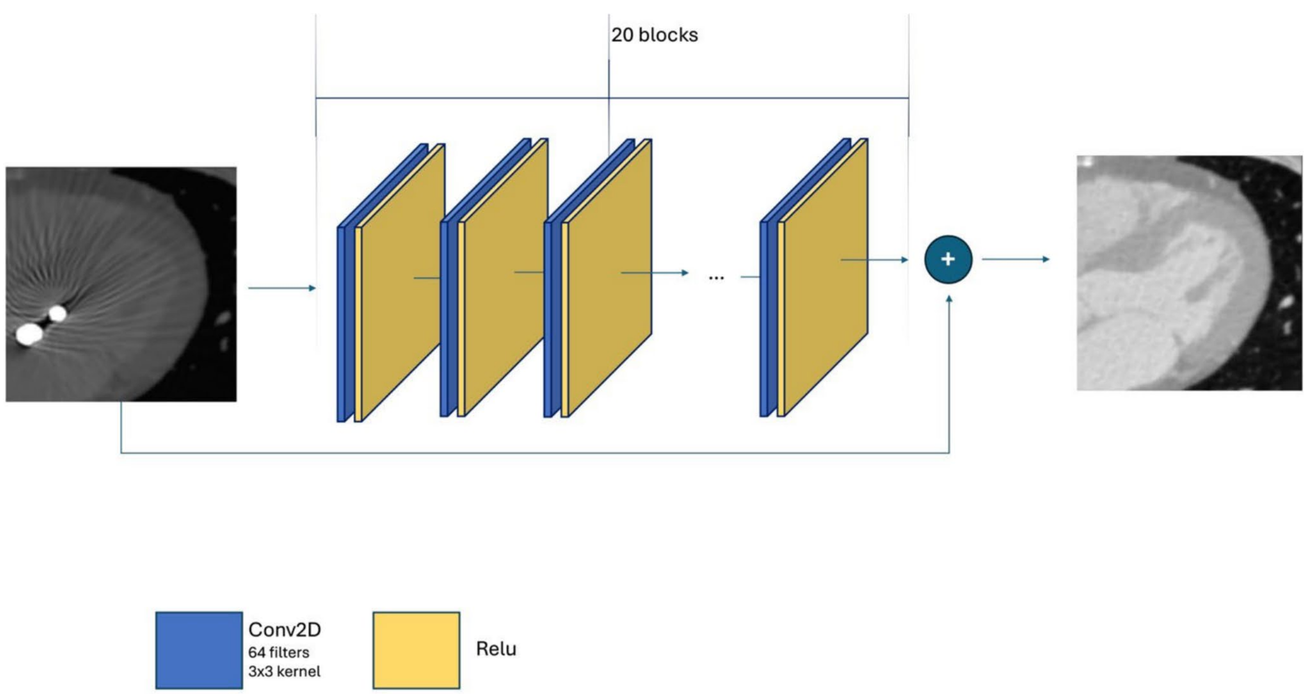

Fig. 2 Deep learning model's architecture consisting of 20 sequential convolutional layers with 3×3 kernels and ReLU activation functions. The skip connection positioned before the output layer enables concatenation of the input image with final layer features, preserving structural integrity and facilitating gradient flow during network training

bias related to patient selection and assess the robustness of the method, the training procedure was repeated 10 times, each time randomly selecting two patients to form the test set. Table 1 reports the distribution of patients and corresponding images across the training, validation, and test sets over the 10 data splits. The dataset was divided at patient level to ensure independence between sets; as a result, the number of images in each set may vary across the splits. Although the number of patients is the same in each split, the specific included patients differ. Ultimately, one of the ten trained models—the one achieving the highest performance on the validation set—was selected for subsequent analysis. From this point onward, the term *filtered image* denotes the image after being processed by our deep learning model.

Performance Metrics

To comprehensively evaluate model performance, we computed two types of quantitative metrics: global image-based and region-specific metrics focused on the left ventricle (LV). This dual evaluation strategy was adopted to assess both the overall quality of artifact reduction across the entire CCT slice and the accuracy of reconstruction within the clinically relevant region for subsequent radiomic analysis. SSIM, MAE, and mean squared error (MSE) were employed as global metrics. SSIM is an index designed to assess image quality by taking into account the structural properties of an image, with values ranging from -1 to $+1$, where $+1$ indicates identical images. To further investigate the model's performance in the most relevant anatomical region, these metrics were also adapted to the LV region. In particular, SSIM was computed only on the pixels corresponding to the LV wall identified by the segmentation mask, resulting in a dedicated metric termed LV-SSIM. This targeted approach provides a more precise assessment of the network's effectiveness in preserving critical cardiac structures.

Similarly, MAE and MSE were computed within the LV region by comparing the predicted values with the corresponding reference standards, yielding the LV-MAE and LV-MSE metrics. These measures provide a complementary quantitative evaluation of reconstruction accuracy within the LV wall.

To compare the performance of the proposed network with the state of the art, the dataset was also analyzed using two additional neural networks: one commonly adopted in the literature (ACDNet [21]) and a more recent method Cardiac Inpainting Net [14]. The ACDNet is an adaptive convolutional dictionary network for metal artifact reduction in CT images, but not specifically for CCT, and it is constructed by mapping the steps of a model-based optimization algorithm into successive network modules. This design explicitly embeds the structure of metal artifacts into the network, resulting in an interpretable architecture tailored to the MAR task. ACDNet combines interpretability with representation capability since each part of the net corresponds to a specific part of the optimization algorithm, while adaptively learning artifact-free image priors and input-dependent convolutional kernels. The Cardiac Inpainting Net, as the ACDNet, does not rely on raw projection data and employs deep learning to identify and reconstruct metal-corrupted regions directly from reconstructed CT images. Artifact detection and image inpainting are handled by two dedicated networks. In particular, the inpainting network is based on a GAN architecture derived from the Pix2Pix framework [22], which is used to paint the affected regions and so suppress metal-induced artifacts. To compare the proposed network to the ACDNet and the Cardiac Inpainting Net, a paired Wilcoxon test is performed (statistical significance was defined as p-value < 0.05).

When the proposed network is applied to IVF patient data, due to the nature of the problem, ground-truth references are not available; therefore, only visual assessment can be performed. A qualitative artifact score, called patient artifact index, was assigned to each patient scan by three expert radiologists before and after MAR algorithm application. This index goes from 5 to 1: lower scores correspond to optimal quality images with minimal artifacts and well-defined cardiac structures, whereas higher scores indicate severe artifact contamination, with prominent streak artifacts and reduced anatomical interpretability.

Radiomics

A total of 851 3D radiomic features were extracted from the LV wall's volume. In particular, 14 features pertain to the shape and size class, 18 features to the first-order statistics category, 75 features capture textural characteristics based on grey level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM), and the remaining 744 features were extracted from wavelet transformations of the original image. Radiomic features were extracted from filtered and artifact-free images belonging to the control group and from filtered images of IVF patients. Default Pyradiomics 3.0 settings were considered for extracting 3D radiomic features, namely, B-spline

Table 1 Average distribution of patients and corresponding images across training, validation, and test sets over the 10 splits

	Training	Validation	Test
Patients	10	1	2
Images	2368 ± 164	271 ± 80	500 ± 167

interpolation and fixed-bin histogram discretization, with 25 bins. As the features were obtained using Pyradiomics, they were compatible with the Image Biomarker Standardization Initiative.

Radiomic Feature Assessment

To further evaluate the artifact removal method, features extracted from the original artifact-free images and those obtained from the filtered images were compared. This assessment quantified the network's ability to remove noise while preserving essential image characteristics, quantified by radiomics. The intraclass correlation coefficient (ICC) was computed to measure feature similarity. Following the established guidelines [23], an ICC threshold of 0.75 was set and only features exceeding this threshold were considered similar.

Another critical step of the radiomic workflow is the stability analysis that examines feature sensitivity to variations in region of interest (ROI) delineation, which typically occur during manual segmentation. This step was performed on the artifact-free images. To simulate human segmentation variability, three small-entity perturbations, namely, erosion, dilation, and contour randomization, were applied to the original ROIs of control group's scans, as in Lo Iacono et al. [16]. The ICC was calculated to quantify agreement between features from the original ROI and the ROI after perturbation and the same threshold set before was employed to identify stable features.

When considering IVF patients, only features both robust to segmentation variations and well-preserved by the filtering process were retained for subsequent analysis.

Machine Learning Model Development

Features selected in the assessment described in the previous paragraph were employed to develop a classification model aimed at predicting a composite endpoint in IVF patients.

Given the very small size of the dataset, a fourfold nested cross-validation (Nested CV) scheme was implemented. This approach employs two nested levels of fourfold cross-validation as part of the validation process: an inner loop for model selection, including the choice of both the learning algorithm and the feature selection strategy, and an outer loop for providing an unbiased estimate of generalization performance. This separation ensures that the final evaluation is not influenced by decisions made during the model selection phase, thereby reducing the risk of overfitting and overly optimistic performance estimates.

In each Nested-CV iteration, data balancing was applied, and various feature selection methods and classification algorithms were systematically evaluated. As the dataset exhibited class imbalance, with positive samples ($n=23$) outnumbering negative samples ($n=8$), balancing was performed by oversampling the minority class, employing ROI image perturbations (erosion, dilation, and contour randomization) as described in Lo Iacono et al. [16].

Feature selection comprised two sequential steps: nonredundancy and relevance analysis. The nonredundancy analysis calculated the Spearman correlation coefficient (ρ) between feature pairs: when $|\rho|$ exceeded a predefined threshold, only the feature with lower mean correlation relative to all other features was retained. The relevance analysis included five different feature selection methods: statistical analysis using Mann–Whitney test, least absolute shrinkage and selection operator (LASSO), semisupervised LASSO (ssLASSO), principal component analysis (PCA), and semisupervised PCA (ssPCA).

Each feature selection method was tested in combination with four classifiers including decision tree (DT), logistic regression (LR), extreme gradient boosting (XGB), and random forest (RF). Model performance was evaluated using the $F1$ score (defined as the harmonic mean of precision and recall), balanced accuracy (defined as the average of sensitivity and specificity), and sensitivity and specificity. The combination providing the highest average $F1$ score on validation data was selected as the optimal pipeline.

Results

DL-Based Metal Artifact Reduction

Across ten independent experiments, the average performance on the validation sets was an MAE of 25.80 ± 2.04, an MSE of 1560.87 ± 243.94, an SSIM of 0.908 ± 0.011, and a PSNR of 33.25 ± 0.79, see Supplementary Material Table 1. The corresponding average performance on the test sets yielded an MAE of 26.74 ± 2.36 HU, an MSE of 1680.91 ± 270.94 HU, an SSIM of 0.906 ± 0.014, and a PSNR of 33.11 ± 1.02 dB. The close agreement between validation and test results indicates that the proposed method exhibits robust and stable performance, independent of the specific test set selection.

Across the ten experiments, the network configuration with the highest validation performance, selected for further analysis, achieved a mean MAE of 24.19 ± 6.90 HU, an MSE of 1378.41 ± 788.14 HU, a PSNR of 35.14 ± 1.87 dB, and an SSIM of 0.932 ± 0.026 on the validation set. On its corresponding test set, the model obtained a mean MAE of 24.85 ± 7.16 HU, an MSE of 1439.62 ± 883.72 HU, a PSNR of 33.47 ± 1.96 dB, and an SSIM of 0.911 ± 0.034. Table 2 shows the LV-MAE, LV-MSE, and LV-SSIM values for individual patients in the training and test sets, respectively. Examples from two patients are shown in Fig. 3, where the filtered images are compared to the ground truth images.

To further contextualize the effectiveness of the proposed solution, we compared its performance against two alternative state-of-the-art networks: Cardiac Inpainting Net and

Table 2 Network performance in terms of mean absolute error (LV-MAE), mean squared error (LV-MSE), and structural similarity index (LV-SSIM), computed on subjects belonging to the training and test sets for the model selected among the ten independent splits

	LV-MAE (HU)	LV-MSE (HU)	LV-SSIM
Training set			
Subject 1	17.64 ± 8.45	670.19 ± 901.29	0.915 ± 0.050
Subject 2	14.60 ± 5.88	463.08 ± 522.92	0.948 ± 0.034
Subject 3	14.95 ± 5.39	449.27 ± 452.20	0.943 ± 0.034
Subject 4	16.36 ± 6.49	557.22 ± 684.73	0.935 ± 0.036
Subject 5	15.57 ± 5.79	484.57 ± 383.97	0.917 ± 0.049
Subject 6	16.15 ± 6.65	568.77 ± 647.67	0.933 ± 0.041
Subject 7	13.98 ± 5.80	436.22 ± 641.83	0.939 ± 0.035
Subject 8	14.29 ± 6.51	448.63 ± 596.13	0.949 ± 0.037
Subject 9	16.48 ± 6.65	552.75 ± 629.36	0.928 ± 0.038
Subject 10	17.31 ± 4.16	520.27 ± 304.38	0.922 ± 0.025
Subject 11	15.64 ± 8.53	642.16 ± 1202.14	0.939 ± 0.042
Average	15.72 ± 6.39	526.65 ± 633.33	0.934 ± 0.038
Test set			
Subject 12	15.66 ± 9.38	656.80 ± 1129.51	0.934 ± 0.048
Subject 13	15.52 ± 7.59	566.99 ± 878.09	0.939 ± 0.043
Average	15.59 ± 8.48	661.89 ± 1003.80	0.936 ± 0.045

ACDNet. As shown by the performance metrics shown in Table 3, the three methods exhibit markedly different performance levels. Cardiac Inpainting Net achieved an SSIM of 0.626 ± 0.034, ACDNet yielded 0.773 ± 0.049, whereas the proposed network attained a significantly higher SSIM of 0.902 ± 0.038 ($p < 0.001$). Supplementary Fig. 1 shows a comparison between filtered images and results by Cardiac Inpainting Net and ACDNet.

The proposed network was then applied on IVF patient scans. The average patient artifact index of the IVF patient was 3.6 (median = 4), while after MAR the average score became 1.6 (median = 1), see Supplementary Table 2.

Figure 4 presents representative examples of images processed by the network. For five IVF patients who initially received a patient artifact index of 5, indicating severe artifacts, DL-based filtering improved the patient artifact index up to 3. Despite this improvement, a moderate level of residual artifacts remained, which did not satisfy the predefined criteria for reliable radiomic analysis; consequently, these cases were excluded from further evaluation.

Radiomics

Radiomic Feature Assessment

Feature assessment was performed using the control group data. The similarity analysis compared radiomic features calculated from the original artifact-free CCT images with the corresponding five filtered images. This analysis revealed 779 similar features (92% of the 851 initial features), demonstrating the network's ability to preserve image characteristics during artifact removal.

The stability analysis, performed on the original CCT images from the control group by applying ROI perturbations (erosion, dilation, and contour randomization), identified 433 stable features (51%) out of the 851 total extracted features.

Thus, when considering only stable features, 424 radiomic features were also identified as similar after artifact removal, i.e., 98% of all the stable features. Table 4 reports a summary of feature analysis results, highlighting which features exhibit similarity, stability, or both properties, organized according to their radiomic class.

Composite Endpoint Prediction in IVF Patients

To identify the optimal combination of feature selection method and classification model, performance was evaluated within the inner loop of the nested cross-validation. The highest performance was achieved by combining ssPCA with the XGB classifier. This configuration was subsequently evaluated across all testing folds of the outer loop.

Test results averaged over the outer cross-validation folds are presented in Table 5. It can be noted that an F1 score of 0.85 ± 0.07 is reached. The corresponding confusion matrix, obtained by aggregating predictions across the outer folds, is shown in Fig. 5.

Discussion

As CCT continues to establish its role in cardiovascular assessment and can be used as an alternative when CMR is not possible anymore due to ICD, improving image quality for patients with metallic implants remains essential for ensuring comprehensive diagnostic evaluation across diverse patient populations. The current study addresses a critical challenge in this framework by developing a DL approach that works directly on CT slices to reduce metal artifacts caused by ICDs. By leveraging a CNN and creating a simulated dataset, this study demonstrated effective reduction of metallic artifacts in CCT images, with an average SSIM equal to 0.902 ± 0.039 compared to 0.773 ± 0.05 and 0.624 ± 0.03 obtained by respectively the ACDNet and the Cardiac Inpainting Net on the same dataset. The radiomic analysis validated the model's effectiveness, showing that 98% of features extracted from filtered images were statistically similar to the stable features extracted from original artifact-free images, confirming preservation of critical quantitative information. This filtering technique ensures that the reconstructed images are not only visually improved

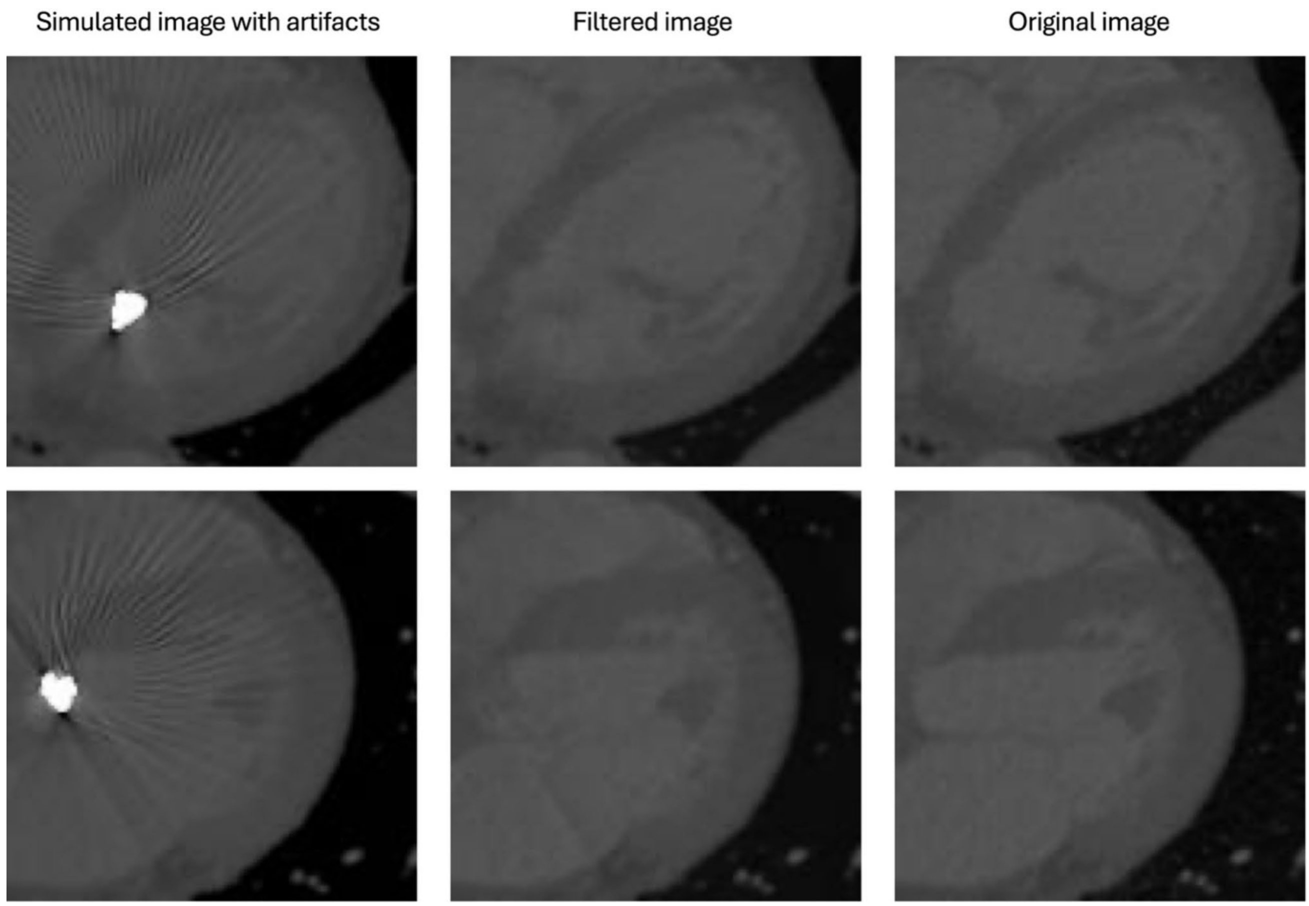

Fig. 3 Examples of artifact removal results. The left column displays simulated images with artifacts, the middle column shows the results after artifact reduction using the proposed model, and the right column presents the original artifact-free images for comparison. The SSIM for the two images is 0.93 (top row) and 0.90 (bottom row)

Table 3 Comparison of test set performance for three neural networks—ACDNet, Cardiac Inpainting Net, and the network proposed in this paper—evaluated using mean absolute error (MAE), mean squared error (MSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR). Error-based metrics (MAE and MSE) are reported in Hounsfield units (HU). In addition, left ventricular–specific metrics, including LV-MAE, LV-MSE, and LV-SSIM, are reported to assess reconstruction quality within the LV wall region. The last column reports the *p*-values derived from the Wilcoxon signed-rank test, used to assess statistically significant differences between methods

	ACDNet	Cardiac Inpainting Net	Proposed network
MAE (HU)	78.20 ± 13.44	87.60 ± 8.07	20.47 ± 4.10 *°
MSE (HU)	20,943.49 ± 22,390.26	15,343.35 ± 2551.24	1240.90 ± 754.60 *°
SSIM	0.773 ± 0.05	0.624 ± 0.03	0.902 ± 0.039 *°
PSNR (dB)	21.49 ± 2.75	22.07 ± 0.61	33.12 ± 2.21 *°
LV-MAE (HU)	39.38 ± 3.82	40.94 ± 19.90	15.67 ± 6.97 *°
LV-MSE (HU)	12,592.38 ± 4603.05	4600.88 ± 5977.45	559.91 ± 673.19 *°
LV-SSIM	0.798 ± 0.03	0.699 ± 0.14	0.936 ± 0.04*°

*$p < 0.001$ proposed network vs. ACDnet

°$p < 0.001$ proposed network vs. Cardiac Inpainting Net

but also diagnostically reliable for advanced analysis, particularly in the LV region. This enhancement in image quality was leveraged to address a significant diagnostic challenge in cardiac imaging for IVF patients with defibrillators: in fact, the metallic noise generated in CCT images of these patients typically prevents radiomic analysis of cardiac structures. After

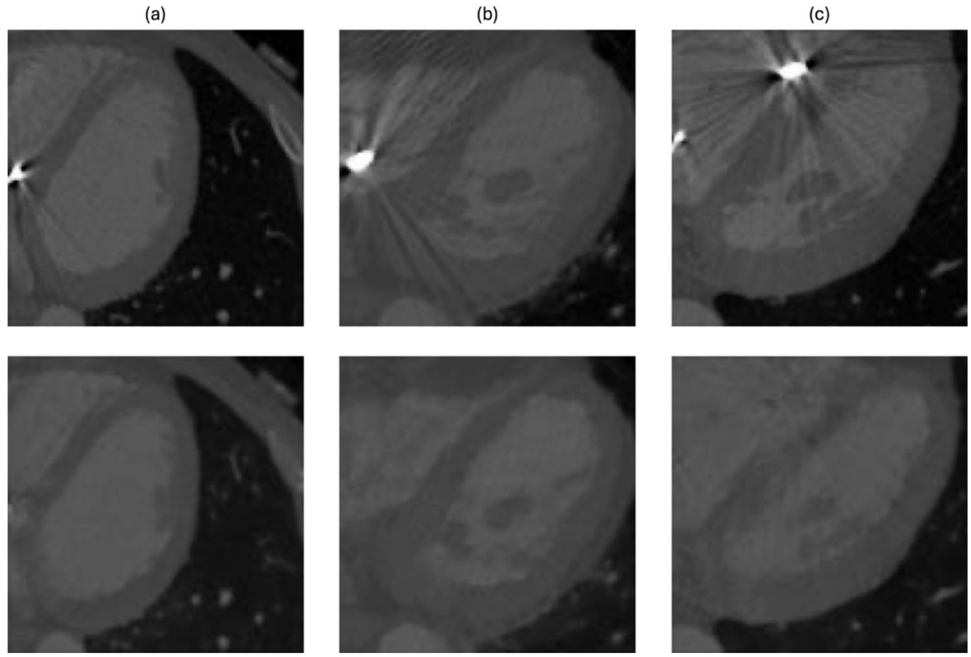

Fig. 4 Examples of artifact reduction in real clinical CCT images from patients with idiopathic ventricular fibrillation. Panels (**a**–**c**) correspond to three slices of three different patients. The top row shows an original slice with metal artifacts, while the bottom row shows the filtered image by the proposed network. The patient artifact index changed from 4 to 1 (**a**), 5 to 1 (**b**), and 4 to 2 (**c**)

Table 4 Number of features identified as similar, stable and fulfilling both criteria. Features are further organized by their class membership, i.e., shape and size, first-order statistics, texture, and wavelet

	First-order statistics	Texture	Wavelet
Similarity (artifact-free vs. filtered images)	18	75	672
Stability (artifact-free vs. perturbed artifact-free)	3	17	403
Similarity and stability	3	17	394

Table 5 Performance metrics of the classification model evaluated using nested cross-validation. Results report the mean and the standard deviation over the outer folds

Metric	
Balanced accuracy	0.78 ± 0.09
Sensibility	0.81 ± 0.14
Specificity	0.75 ± 0.25
F1 score	0.85 ± 0.07

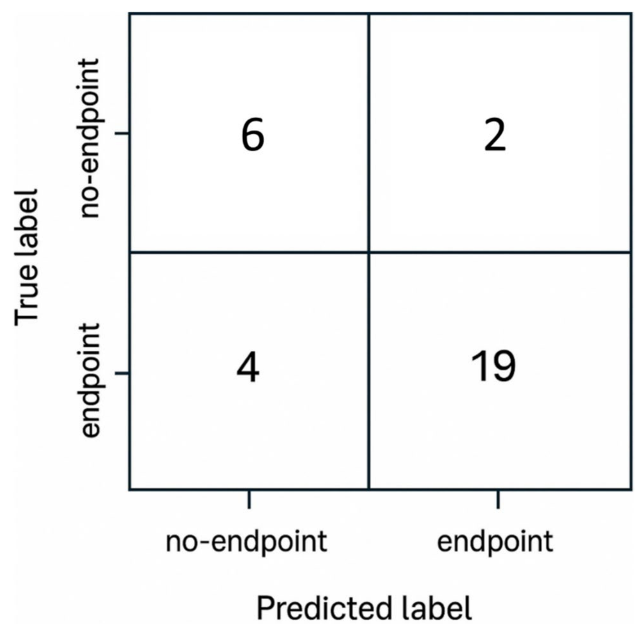

Fig. 5 Confusion matrix showing the predictive performance of the best model configuration (semisupervised principal component analysis and extreme gradient boosting) on the test dataset

applying the developed filtering technique, reliable radiomic features were extracted and used to predict a composite clinical endpoint including NSVT, high PVC burden, and arrhythmia. This predictive model achieved promising performance metrics with an $F1$ score of 0.83, demonstrating the diagnostic utility of the filtered images in clinical applications.

Several approaches have been proposed in the literature to address metal artifacts in CT imaging, but applications specifically targeting cardiac imaging remain limited [3, 8–10, 13, 21, 22]. In this field, some studies addressed MAR by using acquisition improvement strategies like dual-energy protocols for material decomposition, which require specific modifications to the standard CT acquisition protocols and scanner hardware [22, 23]. These techniques have been tested across different metallic implants, with consistent findings regarding optimal energy levels and device-specific considerations. Dual-energy CT with virtual monoenergetic imaging (VMI) has shown significant artifact reduction across various cardiac devices. Secchi et al. [24] demonstrated effective reduction of artifacts from sternal wires, bypass clips, and coronary stents in 35 patients, evaluating both quantitative artifact size reduction (measured in mm) and qualitative image quality scores on a 5-point Likert scale. Schwartz et al. [25] reached an artifact reduction up to 17.2% in CCT images acquired from 80 patients with aortic valve replacements, assessing the percentage of valve area obscured by artifact and measuring opacification (HU values). Their findings showed that a high-keV VMIs achieved a better artifact reduction compared to standard acquisitions, with notable differences between valve materials, highlighting the need for component-specific optimization strategies. Beyond dual-energy approaches, cardiovascular-based literature reports some studies employing projection-based algorithms. Groves et al. [3] investigated the SEMAR algorithm in 122 CCT studies of patients with implanted metal and in a defibrillator lead phantom. Using beam hardening artifact radius and attenuation variation as performance metrics, they found that SEMAR reduced the maximum artifact radius (mm) by 77% and artifact attenuation variation (HU) by 51%. Subjective image quality improved significantly on a 4-point scale (from 3 to 2), with SEMAR reducing the number of nondiagnostic coronary segments by 56%. Similarly, Tatsugami et al. [9] evaluated the SEMAR in 8 patients with cardiac devices, assessing coronary artery visualization on a 4-point scale. While limited in sample size, their results showed improvement on image artifacts, with 81.8% of previously nondiagnostic segments (9 of 11) becoming diagnostic after SEMAR application. In assessing lead perforation, Kidoh et al. [8] evaluated 47 patients with CIEDs and used the SEMAR algorithm in CCT images. They quantified artifact reduction using an artifact index (AI), showing significant reduction from 284.6 ± 134.1 to 96.7 ± 40.1 HU ($p < 0.001$). Notably, this MAR approach improved the diagnostic uncertainty in 27.7% of cases who could not be definitively diagnosed using conventional imaging techniques. MAR algorithms were also employed in combination with VMI reconstruction, as by Pennig et al. [10] who evaluated a comprehensive approach to CIED artifacts in 34 patients by comparing VMI, MAR algorithms, and their combination (VMIMAR). Using spectral detector CT, they found that MAR and VMIMAR significantly improved both hypodense and hyperdense artifacts around device components. Their findings revealed that MAR and VMIMAR significantly improved both hypodense and hyperdense artifacts around devices. However, tailored approaches are required on different device components.

Despite these advancements, several limitations persist across studies. High-keV VMI techniques often suffer from reduced soft tissue contrast, which is particularly problematic around CIED leads where maintaining vessel visualization is critical. This trade-off necessitates device-specific protocols, as demonstrated in [10, 22]. Notably, several studies have employed subjective assessments by radiologists—such as Likert scales for artifact severity, diagnostic confidence, and visualization of adjacent structures, while objective benchmarks and metrics are needed to perform comparisons among the available studies. Additionally, many reconstruction algorithms remain vendor-specific, limiting widespread application, and caution should be exercised when applying MAR algorithms as they may introduce secondary artifacts [5].

Dealing with cardiovascular images, to the best of authors' knowledge, literature reports just two DL-based studies. The less recent one, by Lossau et al. [13], proposed a pacemaker artifact reduction (DyPAR+) automatic approach to reduce streak-shaped metal artifacts caused by moving pacemakers. A three CNN ensemble working directly in the projection domain was built to detect metal shadows in projection data, replace affected line integrals, and model metal locations. A synthetic dataset was built by introducing moving pacemakers leads into 14 clinical cases without pacemaker, to train the network in a supervised way. Model's performance was evaluated on 9 CCT cases with real pacemakers and 2 without, reaching a mean observer ranking of 1.0 (on a 1–3 scale, where 1 represents the highest quality) and a reconstruction MAE of 11.54 ± 2.49 HU. Differently from the current study, MAE was computed on the entire image rather than on a specific ROI. The method showed some limitations with residual streak artifacts due to inconsistencies among 2D projections after inpainting, partial angle artifacts from incomplete metal shadow segmentation, and occasional introduction of new artifacts in neighboring anatomy. In addition, compared to Lossau's approach, our method includes an artifact simulation incorporating material-specific properties, polychromatic projections, and quantum noise for greater clinical realism. Analyzing the most recent work proposed by McKeown et al., the authors introduced a GAN-based approach aimed at inpainting regions affected by metal artifacts to improve segmentation of heart chambers. In their study, synthetic data were generated from artifact-free images by artificially inserting both the metal source and the associated artifacts. The network was trained to reduce these artifacts and evaluated on 148 cases. Within this framework, it is important to emphasize that the Cardiac Inpainting Network represents an effective solution, particularly for enhancing downstream tasks such as cardiac segmentation. However, radiomic analysis entails different requirements, as it depends on the preservation of consistent and physiologically meaningful texture patterns across the image. Inpainting-based approaches, while visually plausible, may introduce local intensity inconsistencies or sharp transitions between reconstructed and original regions, potentially compromising the stability of radiomic features. Visual inspection frequently reveals noticeable discontinuities between inpainted regions and the surrounding anatomical structures; these abrupt intensity shifts and texture inconsistencies can substantially impact both first- and higher-order feature distributions.

For this reason, the proposed method is specifically designed to reduce metal artifacts while minimizing the introduction of nonphysiological texture patterns and preserving anatomical consistency, with particular attention on the left ventricular region.

Certain limitations must be acknowledged. While the current study employed simulated artifacts, extracted from actual CT images, thus mimicking real-world scenarios, their positioning on artifact-free CT images may not always correspond to realistic implant locations encountered in the clinical practice. Additionally, the dataset size, though sufficient for the current implementation, could benefit from expansion to enhance generalizability across different scanner types and implant configurations.

Future work should focus on validating this approach on a larger cohort of patients with real implanted devices and investigating the application of this method to other cardiovascular implants such as stents or valves.

In conclusion, this study should be considered a preliminary investigation due to the limited cohort size and presents a promising DL solution for metal artifact reduction in CCT imaging of patients with ICD. The demonstrated preservation of radiomic features suggests that the approach not only improves image quality but also maintains diagnostic integrity, potentially expanding the utility of CCT in patients with cardiac devices who previously might have been excluded from certain studies.

Differently from many previous approaches that operate on sinogram data, the proposed method works directly on reconstructed CT slices. The robustness of this slice-based technique was enabled by the development of a comprehensive synthetic data generation pipeline. This involved incorporating realistic physical modeling, including polychromatic X-ray interactions, material-dependent attenuation properties, and simulated quantum noise, alongside strategic data diversification techniques such as varied metal mask geometries and controlled boundary modifications, ensuring the training data closely mirrored complex clinical artifact patterns. Finally, the effectiveness of this approach was validated through objective radiomic feature analysis rather than relying solely on subjective visual assessment or conventional image quality metrics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10278-026-02030-x.

Author Contribution NB: formal analysis; investigation; methodology; software; validation; visualization; writing—original draft; and writing—review and editing; FL: conceptualization; formal analysis; investigation; methodology; validation; visualization; writing—original draft; and writing—review and editing; LMV: data curation and writing—review and editing; MG: data curation and writing—review and editing; GP: data curation and writing—review and editing; RJH: data curation and writing—review and editing; VDAC: conceptualization; formal analysis; investigation; supervision; visualization; writing—original draft; and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement. This work was partially funded by the Dutch Heart Foundation (CVON2017-13 VIGILANCE) and by PNRR-MCNT2-2023–12376978–CUP B63C24000460001-PNRR M6C2I2.1 NEXTGENERATION EU.

Declarations

Ethics Approval This study was performed in line with the principles of the Declaration of Helsinki. All procedures were performed in compliance with relevant laws and institutional guidelines and were approved by local institutional ethics review boards (NL67079.068.18/METC18-038).

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication The authors affirm that human research participants provided informed consent for publication of the images.

Competing interests The authors declare no competing interests.

References

1. Haïssaguerre M, Duchateau J, Dubois R, Hocini M, Cheniti G, Sacher F, Lavergne T, Probst V, Surget E, Vigmond E, Welte N, Chauvel R, Derval N, Pambrun T, Jaïs P, Nademanee W, Bernus O: Idiopathic ventricular fibrillation. JACC Clin Electrophysiol 6:591–608, 2020
2. Almahameed ST, Kaufman ES: Idiopathic ventricular fibrillation: diagnosis, ablation of triggers, gaps in knowledge, and future directions. J Innov Card Rhythm Manag 11:4135–4146, 2020
3. Groves DW, Acharya T, Steveson C, Schuzer JL, Rollison SF, Nelson EA, Sirajuddin A, Sathya B, Bronson K, Shanbhag SM, Chen MY: Performance of single-energy metal artifact reduction in cardiac computed tomography: a clinical and phantom study. J Cardiovasc Comput Tomogr 14:510–515, 2020
4. Gjesteby L, De Man B, Jin Y, Paganetti H, Verburg J, Giantsoudi D, Wang G: Metal artifact reduction in CT: where are we after four decades? IEEE Access 4:5826–5849, 2016
5. Selles M, van Osch JAC, Maas M, Boomsma MF, Wellenberg RHH: Advances in metal artifact reduction in CT images: a review

of traditional and novel metal artifact reduction techniques. Eur J Radiol 170:111276–111276, 2024
6. 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 38:450–461, 2018
7. Kalender WA, Hebel R, Ebersberger J: Reduction of CT artifacts caused by metallic implants. Radiology 164:576–577, 1987
8. Kidoh M, Oda S, Nakato K, Sakabe D, Kanazawa H, Takashio S, Nakaura T, Nagayama Y, Sasao A, Hatemura M, Funama Y, Kaikita K, Tsujita K, Ikeda O, Azuma M, Hirai T: Assessment of cardiac implantable electric device lead perforation using a metal artifact reduction algorithm in cardiac computed tomography. Eur J Radiol 136:109530–109530, 2021
9. Tatsugami F, Higaki T, Sakane H, Fukumoto W, Iida M, Baba Y, Fujioka C, Kihara Y, Tsushima S, Awai K: Coronary CT angiography in patients with implanted cardiac devices: initial experience with the metal artefact reduction technique. Br J Radiol 89:20160493–20160493, 2016
10. Pennig L, Zopfs D, Gertz R, Bremm J, Zaeske C, Große Hokamp N, Celik E, Goertz L, Langenbach M, Persigehl T, Gupta A, Borggrefe J, Lennartz S, Laukamp KR: Reduction of CT artifacts from cardiac implantable electronic devices using a combination of virtual monoenergetic images and post-processing algorithms. Eur Radiol 31:7151–7161, 2021
11. Leipsic J, LaBounty TM, Heilbron B, Min JK, Mancini GBJ, Lin FY, Taylor C, Dunning A, Earls JP: Adaptive statistical iterative reconstruction: assessment of image noise and image quality in coronary CT angiography. AJR Am J Roentgenol 195:649–654, 2010
12. Renker M, Ramachandra A, Schoepf UJ, Raupach R, Apfaltrer P, Rowe GW, Vogt S, Flohr TG, Kerl JM, Bauer RW, Fink C, Henzler T: Iterative image reconstruction techniques: applications for cardiac CT. J Cardiovasc Comput Tomogr 5:225–230, 2011
13. Lossau T, Nickisch H, Wissel T, Morlock M, Grass M: Learning metal artifact reduction in cardiac CT images with moving pacemakers. Med Image Anal 61:101655–101655, 2020
14. McKeown T, Gach HM, Hao Y, An H, Robinson CG, Cuculich PS, Yang D: ICD lead and primary metal artifact detection and inpainting in cardiac CT images. Med Phys 52:e17947–e17947, 2025
15. Prell D, Kalender WA, Kyriakou Y: Development, implementation and evaluation of a dedicated metal artefact reduction method for interventional flat-detector CT. Br J Radiol 83:1052–1062, 2010
16. Lo Iacono F, Maragna R, Pontone G, Corino VDA: A novel data augmentation method for radiomics analysis using image perturbations. J Imaging Inform Med 37:2401–2414, 2024
17. Sakamoto M, Hiasa Y, Otake Y, Takao M, Suzuki Y, Sugano N, Sato Y: Automated segmentation of hip and thigh muscles in metal artifact-contaminated CT using convolutional neural network-enhanced normalized metal artifact reduction. arXiv, https://doi.org/10.48550/arXiv.1906.11484, 2019
18. Kyriakou Y, Meyer E, Prell D, Kachelrieß M: Empirical beam hardening correction (EBHC) for CT. Med Phys 37:5179–5187, 2010
19. National Institute of Standards and Technology: XCOM photon cross sections database. Available at https://www.nist.gov/pml/xcom-photon-cross-sections-database. Accessed 22 April 2025
20. Kim J, Lee JK, Lee KM: Accurate image super-resolution using very deep convolutional networks. arXiv, https://doi.org/10.48550/arXiv.1511.04587, 2016
21. Wang H, Li Y, Meng D, Zheng Y: Adaptive convolutional dictionary network for CT metal artifact reduction. arXiv, https://doi.org/10.48550/arXiv.2205.07471, 2022
22. Isola P, Zhu JY, Zhou T, Efros AA: Image-to-image translation with conditional adversarial networks. Proc IEEE Conf Comput Vis Pattern Recognit: 1125–1134, 2017
23. Koo TK, Li MY: A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 15:155–163, 2016
24. Secchi F, De Cecco CN, Spearman JV, Silverman JR, Ebersberger U, Sardanelli F, Schoepf UJ: Monoenergetic extrapolation of cardiac dual energy CT for artifact reduction. Acta Radiol 56:413–418, 2015
25. Schwartz FR, Tailor T, Gaca JG, Kiefer T, Harrison K, Hughes GC, Ramirez-Giraldo JC, Marin D, Hurwitz LM: Impact of dual energy cardiac CT for metal artefact reduction post aortic valve replacement. Eur J Radiol 129:109135–109135, 2020

Authors and Affiliations

Nicholas Benigni[1] · Francesca Lo Iacono[2] · Lisa M. Verheul[3] · Marco Guglielmo[3] · Paul Volders[3] · Rachel ter Bekke[3] · Gianluca Pontone[4,5] · Rutger J. Hassink[6] · Valentina D. A. Corino[1,2]

✉ Valentina D. A. Corino
valentina.corino@polimi.it

Nicholas Benigni
nicholas.benigni@cardiologicomonzino.it

1 CardioTechLab, Centro Cardiologico Monzino IRCCS, Milan, Italy

2 Department of Electronics, Information and Bioengineering, Politecnico Di Milano Milan, Milan, Italy

3 Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands

4 Department of Biomedical, Surgical and Dental Sciences, University of Milan, Milan, Italy

5 Perioperative Cardiology and Cardiovascular Imaging Department, Centro Cardiologico Monzino IRCCS, Milan, Italy

6 Department of Cardiology, Cardiovascular Research Institute Maastricht (CARIM), Maastricht University Medical Center+, Maastricht, The Netherlands