Radiology

ORIGINAL RESEARCH • CARDIAC IMAGING

AI Cardiac MRI Scar Analysis Aids Prediction of Major Arrhythmic Events in the Multicenter DERIVATE Registry

Fahime Ghanbari, MD • Thomas Joyce, PhD • Valentina Lorenzoni, PhD • Andrea I. Guaricci, MD • Anna-Giulia Pavon, MD • Laura Fusini, MD • Daniele Andreini, MD • Mark G. Rabbat, MD • Giovanni Donato Aquaro, MD • Raffaele Abete, MD • Jan Bogaert, MD, PhD • Giovanni Camastra, MD • Samuela Carigi, MD • Nazario Carrabba, MD • Grazia Casavecchia, MD • Stefano Censi, MD • Gloria Cicala, MD • Carlo N. De Cecco, MD, PhD • Manuel De Lazzari, MD • Gabriella Di Giovine, MD • Mauro Di Roma, MD • Marta Focardi, MD • Nicola Gaibazzi, MD, PhD • Annalaura Gismondi, MD • Matteo Gravina, MD • Chiara Lanzillo, MD, PhD • Massimo Lombardi, MD, PhD • Jordi Lozano-Torres, MD • Ambra Masi, MD • Claudio Moro, MD • Giuseppe Muscogiuri, MD, PhD • Alberto Nese, MD • Silvia Pradella, MD • Stefano Sbarbati, MD • U. Joseph Schoepf, MD • Adele Valentini, MD • Gérard Crelier, PhD • Pier Giorgio Masci, MD, PhD • Gianluca Pontone, MD • Sebastian Kozerke, PhD • Juerg Schwitter, MD

From the Cardiovascular Department, CMR Center, University Hospital Lausanne–CHUV, Rue du Bugnon 46, 1011 Lausanne, Switzerland (F.G., A.G.P., J.S.); Faculty of Biology and Medicine, Lausanne University, UniL, Lausanne, Switzerland (F.G., J.S.); Institute for Biomedical Engineering, University and ETH Zurich, Zurich, Switzerland (T.J., S.K.); Institute of Management, Scuola Superiore Sant'Anna, Pisa, Italy (V.L.); Institute of Cardiovascular Disease, Department of Emergency and Organ Transplantation, University Hospital Policlinico of Bari, Bari, Italy (A.I.G.); Centro Cardiologico Monzino IRCCS, Milan, Italy (L.F., D.A., G.P.); Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy (L.F.); Loyola University of Chicago, Chicago, Ill (M.G.R.); Edward Hines Jr. VA Hospital, Hines, Ill (M.G.R.); U.O.C. Risonanza Magnetica per Immagini, Fondazione G. Monasterio CNR-Regione Toscana, Pisa, Italy (G.D.A.); Department of Cardiology, Policlinico di Monza, Monza, Italy (R.A., G.D.G.); Department of Radiology, University Hospital Leuven, Leuven, Belgium (J.B.); Cardiac Department, Vannini Hospital Rome, Rome, Italy (G. Camastra); Department of Cardiology, Infermi Hospital, Rimini, Italy (S. Carigi); Cardiovascular and Thoracic Department, Careggi Hospital, Florence, Italy (N.C.); Department of Medical and Surgical Sciences (G. Casavecchia), Department of Radiology (M.G.), University of Foggia, Foggia, Italy; Maria Cecilia Hospital, GVM Care & Research, Cotignola, Italy (S. Censi); Scienze Radiologiche, Department of Medicine and Surgery, University of Parma, Parma, Italy (G. Cicala); Division of Cardiothoracic Imaging, Department of Radiology and Biomedical Informatics, Emory University, Atlanta, Ga (C.N.D.C.); Department of Cardiac, Thoracic, Vascular Sciences and Public Health University of Padua, Medical School, Padova, Italy (M.D.L.); Radiology Department, Policlinico Casilino, Rome, Italy (M.D.R.); Department of Medical Biotechnologies, Division of Cardiology, University of Siena, Siena, Italy (M.F., A.G.); Department of Cardiology, Azienda Ospedaliero-Universitaria, Parma, Italy (N.G.); Cardiology Department, Policlinico Casilino, Rome, Italy (C.L.); Multimodality Cardiac Imaging Section, IRCCS Policlinico San Donato, San Donato Milanese, Milan, Italy (M.L.); Hospital Universitari Vall d'Hebron, Department of Cardiology, Vall d'Hebron Institut de Recerca (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain (J.L.T.); Centro de Investigación Biomédica en Red-CV, CIBER CV, Madrid, Spain (J.L.T.); De Gasperis' Cardio Center, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy (A.M.); Department of Cardiology, ASST Brianza, Desio, Italy (C.M.); Department of Radiology, IRCCS Istituto Auxologico Italiano, San Luca Hospital, Milan, Italy (G.M.); University Milano Bicocca, Milan, Italy (G.M.); Dipartimento Neuro-Cardiovascolare, Ospedale Ca' Foncello Treviso, Treviso, Italy (A.N.); Department of Radiology, Careggi Hospital, Florence, Italy (A.P.); Radiology Department, Vannini Hospital Rome, Rome, Italy (S.S.); Division of Cardiovascular Imaging, Department of Radiology and Radiological Science, Medical University of South Carolina, Charleston, SC (U.J.S.); Department of Radiology, Fondazione IRCCS Policlinico S. Matteo, Pavia, Italy (A.V.); GyroTools, Winterthur, Switzerland (G. Crelier); and School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK (P.G.M.). Received September 9, 2022; revision requested November 1; revision received December 14; accepted February 9, 2023. **Address correspondence to** J.S. (email: *jurg.schwitter@chuv.ch*).

Supported by the Swiss National Science Foundation (SNSF) (grant number 32003B_159727 SNF) (Reagik Substudy) and the Italian Ministry of Health (RC 2017 R659/17-CCM698). S.K. is supported by HRT SWISSHEART Failure Network, Swiss National Science Foundation (grant 325230_197702).

Conflicts of interest are listed at the end of this article.

Radiology 2023; 307(3):e222239 • https://doi.org/10.1148/radiol.222239 • Content codes: CA MR AI

Background: Scar burden with late gadolinium enhancement (LGE) cardiac MRI (CMR) predicts arrhythmic events in patients with postinfarction in single-center studies. However, LGE analysis requires experienced human observers, is time consuming, and introduces variability.

Purpose: To test whether postinfarct scar with LGE CMR can be quantified fully automatically by machines and to compare the ability of LGE CMR scar analyzed by humans and machines to predict arrhythmic events.

Materials and Methods: This study is a retrospective analysis of the multicenter, multivendor CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy (DERIVATE) registry. Patients with chronic heart failure, echocardiographic left ventricular ejection fraction (LVEF) of less than 50%, and LGE CMR were recruited (from January 2015 through December 2020). In the current study, only patients with ischemic cardiomyopathy were included. Quantification of total, dense, and nondense scars was carried out by two experienced readers or a Ternaus network, trained and tested with LGE images of 515 and 246 patients, respectively. Univariable and multivariable Cox analyses were used to assess patient and cardiac characteristics associated with a major adverse cardiac event (MACE). Area under the receiver operating characteristic curve (AUC) was used to compare model performances.

Results: In 761 patients (mean age, 65 years ± 11, 671 men), 83 MACEs occurred. With use of the testing group, univariable Cox-analysis found New York Heart Association class, left ventricle volume and/or function parameters (by echocardiography or CMR), guideline criterion (LVEF of ≤35% and New York Heart Association class II or III), and LGE scar analyzed by humans or the machine-learning algorithm as predictors of MACE. Machine-based dense or total scar conferred incremental value over the guideline criterion for the association with MACE (AUC: 0.68 vs 0.63, $P = .02$ and AUC: 0.67 vs 0.63, $P = .01$, respectively). Modeling with competing risks yielded for dense and total scar (AUC: 0.67 vs 0.61, $P = .01$ and AUC: 0.66 vs 0.61, $P = .005$, respectively).

Conclusion: In this analysis of the multicenter CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy (DERIVATE) registry, fully automatic machine learning–based late gadolinium enhancement analysis reliably quantifies myocardial scar mass and improves the current prediction model that uses guideline-based risk criteria for implantable cardioverter defibrillator implantation.

ClinicalTrials.gov registration no.: NCT03352648

Supplemental material is available for this article.

Abbreviations

AUC = area under the receiver operating characteristic curve, CMR = cardiac MRI, DERIVATE = CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy, ICD = implantable cardioverter defibrillator, LGE = late gadolinium enhancement, LVEF = left ventricular ejection fraction, MACE = major adverse cardiac event, SCD = sudden cardiac death, TTA = test-time augmentation

Summary

A fully automatic machine learning–based analysis of late gadolinium enhancement cardiac MRI quantified myocardial scar and improved prediction of major adverse cardiac events compared with current guidelines.

Key Results

- In a retrospective analysis of 761 patients with chronic heart failure from the multicenter, multivendor CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy (DERIVATE) study, a machine learning algorithm quantified the mass of late gadolinium enhancement cardiac MRI comparable to humans (median absolute error, 2.9 g for dense scar).
- Dense scar quantification by machine learning was superior to prediction of major adverse cardiac events by current guidelines (area under the receiver operating characteristic curve: 0.68 vs 0.63, respectively; $P = .02$).

Cardiac diseases, including myocardial infarction, are associated with a high mortality rate accounting for approximately 30% of all deaths in industrialized countries (1), with sudden cardiac death (SCD) causing approximately half of the postmyocardial infarction deaths (2). In patients with coronary artery disease, myocardial scar mass can be determined precisely with late gadolinium enhancement (LGE) cardiac MRI (CMR) (3,4). Scar burden with LGE CMR has been shown to predict all-cause mortality (5), appropriate implantable cardioverter defibrillator (ICD) shocks (6), SCD, and a combined arrhythmic end point (7). Two meta-analyses also found positive correlations between LGE scar mass and arrhythmic outcomes (8,9). These findings warrant testing in larger multicenter data sets to address whether the prediction of postmyocardial infarction arrhythmic events by LGE scar quantification can be generalized to a broad patient population across centers with different MR scanners, protocols, and personnel. With this in mind, the CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy (DERIVATE) cohort recruited patients from 21 centers in Europe and the United States for this study (10). Current guidelines assign a class 1 indication for ICD implantation in patients with coronary artery disease with left ventricle ejection fraction (LVEF) of 35% or smaller and New York Heart Association class II or III dyspnea (11–13). In absolute numbers, however, about two-thirds of SCDs occur in patients with a LVEF greater than 35% (14), indicating the need to identify additional risk factors for arrhythmic events beyond LVEF and symptoms. Therefore, DERIVATE included patients with LVEF up to 50%.

Currently, LGE scar is quantified by humans, which is time consuming, adds variability to the results, and ultimately limits its generalizability. Recently, artificial intelligence approaches have emerged for MRI analyses (15,16). Thus, in this study, we developed a machine learning algorithm on the DERIVATE LGE image data. We tested whether the algorithm could fully, automatically quantify postinfarct scar with LGE images to yield total, dense, and nondense scar mass and compare the ability of LGE CMR scar analyzed by humans and machines to predict arrhythmic events.

Materials and Methods

Study Participants

In DERIVATE, an international, multicenter, multivendor, observational registry (Clinicaltrials.gov registration no.: NCT

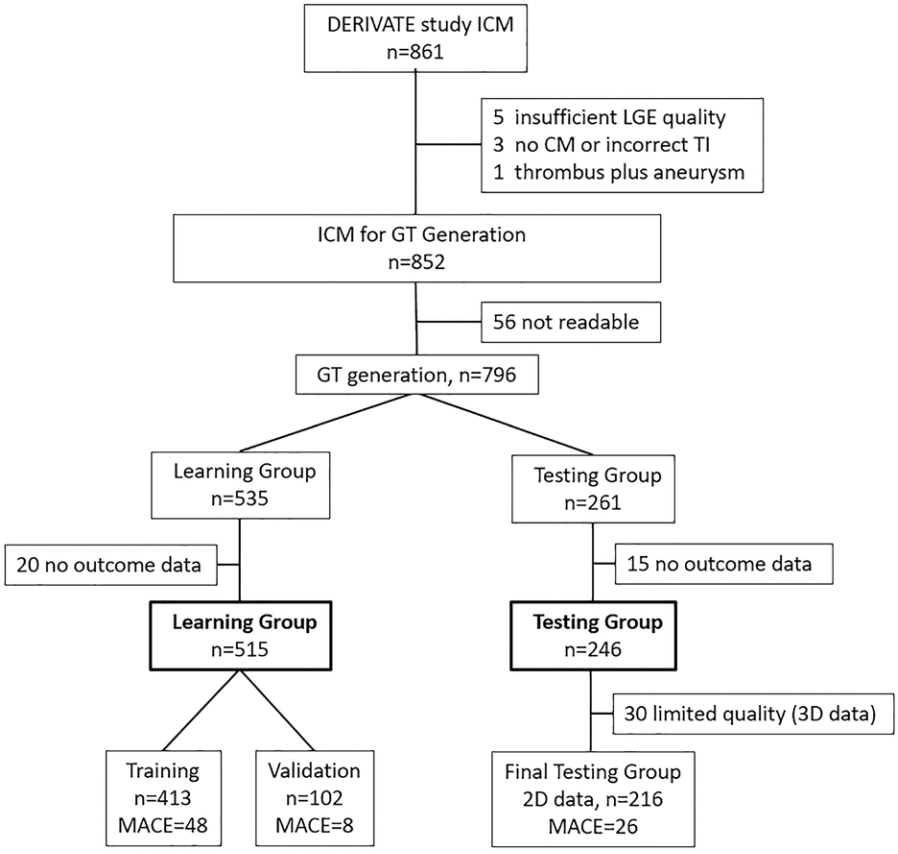

Figure 1: Participant flow diagram shows exclusion criteria and the learning (two-thirds of data set) and testing (one-third of data set) groups used for the machine learning algorithm. CM = contrast medium, DERIVATE = CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy, GT = ground truth, ICM = ischemic cardiomyopathy, LGE = late gadolinium enhancement, MACE = major adverse cardiac event, TI = inversion time, 3D = three dimensional, 2D = two dimensional.

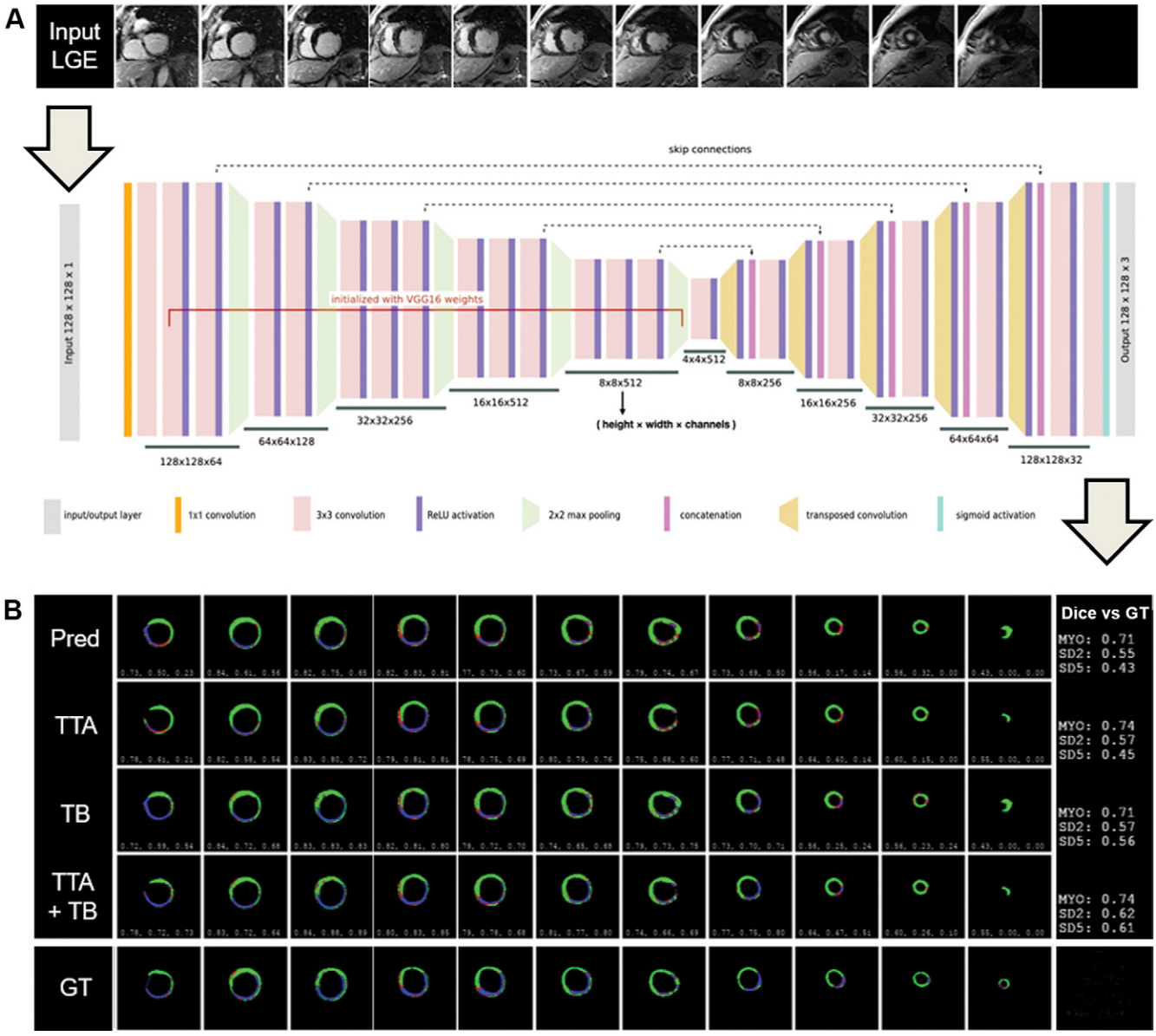

Figure 2: Charts show Ternaus network architecture. **(A)** Late gadolinium enhancement (LGE) images were used as input. Images in **A** cover the left ventricle from base to apex of one patient. The Ternaus network used in this study features a U-Net architecture in which the down-sampling path is initialized with VGG16 weights. As VGG16 expects RGB (ie, three-channel) images, we prepend a single convolutional layer of three 1 × 1 kernels. VGG16 refers to a convolutional neural network with a depth of 16 layers that was pretrained with more than a million images from the ImageNet database *(www.image-net.org)*. ReLu denotes rectified linear units; max pooling calculates the maximum value of the feature maps; skip connections feed the output of layers of the down-sampling path as input to the up-sampling paths without further convolutions; sigmoid activation converts the input to values between 0 and 1 based on a sigmoid function. **(B)** Rows 1–4 show the outputs produced from images in **A** by the machine learning algorithm: prediction (Pred) is the output of the Ternaus network (yielding total left ventricular mass, dense scar, and total scar); test-time augmentation (TTA) predicts the total left ventricular mass, dense scar, and total scar with a set of eight images created of a single LGE image (by rotations, reflections, zooms, and shears) where the eight outputs (after having undergone the reverse rotations, reflections, zooms, and shears) are averaged to yield the final TTA output; the threshold-based (TB) output is the result of the Ternaus network to define normal myocardium (MYO) and the standard reference regions of normal myocardium, from which then dense and total scar are derived by the SD method; and finally, row 4 illustrates the result of a combination of the TTA and the threshold-based approach (TTA+TB). Green represents normal remote myocardium, red represents total-2SD-scar (SD2), and blue represents dense-5SD-scar (SD5). For comparison, the bottom row shows the ground truth (GT) analysis as performed by an experienced observer. Dice versus ground truth refers to the Dice similarity coefficients, where 0 means no overlapping of the machine-generated contours with the ground truth contours and 1 means 100% overlap.

03352648), patients (age, ≥18 years) from 21 centers across Europe and the United States were recruited from January 2015 through December 2020, if they fulfilled the following criteria: *(a)* had established diagnosis of chronic heart failure (17), *(b)* had LVEF of less than 50% with transthoracic echocardiography, and *(c)* had undergone a LGE CMR examination for scar detection. For the presented retrospective analysis, only DERIVATE patients with chronic heart failure of ischemic origin were included (Fig 1, Table S1) (10). Demographics, transthoracic echocardiography, CMR, and outcome data were collected by each local institution up to December 2021. In case of missing data, centers were contacted to revise and complete their database inputs. The study protocol complies with the Declaration of Helsinki and was approved by the lo-

Table 1: Demographics, Medical History, and Clinical Cardiovascular Characteristics of Participants

Characteristic	All Patients (*n* = 761)	Learning Group (*n* = 515)	Testing Group (*n* = 246)	*P* Value*
Age (y)	65 ± 11	65 ± 11	57 ± 12	.47
Sex (M)†	88.2	85.8	93.1	.004
Weight (kg)	79.3 ± 15.3	79.5 ± 16.0	78.8 ± 13.8	.55
Height (m)	1.71 ± 0.08	1.72 ± 0.09	1.71 ± 0.08	.98
BMI (kg/m^2)	26.9 ± 4.4	26.9 ± 4.6	26.7 ± 3.9	.54
BSA (m^2)	1.92 ± 0.20	1.92 ± 0.21	1.91 ± 0.19	.37
NYHA class				
NYHA class I	251 (33)	160 (31)	91 (37)	.10
NYHA class II	314 (41)	212 (41)	102 (42)	.93
NYHA class III	152 (20)	109 (21)	43 (17)	.23
NYHA class IV	44 (6)	34 (7)	10 (4)	.16
Risk factor				
Family history	233 (31)	155 (30)	78 (32)	.65
Hyperlipidemia	470 (62)	318 (62)	153 (62)	.90
Hypertension	501 (66)	335 (65)	165 (67)	.57
Smoking†	341 (45)	246 (48)	96 (39)	.03
Diabetes†	244 (32)	180 (35)	66 (27)	.02
Treatment				
ICD implantation	371 (49)	250 (49)	121 (49)	.86
Betablockers	677 (89)	458 (89)	216 (88)	.82
Ivabradine	61 (8)	46 (9)	17 (7)	.61
ACE/AT1 blockers	632 (83)	422 (82)	209 (85)	.35
Diuretics	556 (73)	386 (75)	170 (69)	.09
Calcium blockers	61 (8)	41 (8)	25 (10)	.28
Antithrombotics	654 (86)	438 (85)	212 (86)	.67
Anticoagulation	198 (26)	134 (26)	62 (25)	.67
Nitrates	122 (16)	88 (17)	34 (14)	.24
Statins	586 (77)	397 (77)	185 (75)	.48
Other antiarrhythmics	53 (7)	36 (7)	17 (7)	.95
Endpoint				
MACE endpoint	83 (11)	56 (11)	27 (13)	.53
SCD	8 (1)	5 (1)	3 (1)	.75
Aborted SCD‡	42 (6)	28 (5)	14 (7)	.88
Sustained VT	52 (7)	33 (6)	19 (9)	.49
All-cause death	94 (12.4)	67 (13)	27 (11)	>.99
Human-determined imaging parameter				
LVEF-echo (%)	34.1 ± 9.8	33.7 ± 9.7	34.9 ± 9.9	.08
LVEF CMR (%)	31.1 ± 9.6	31.1 ± 9.5	31.4 ± 9.9	.60
Total scar (g)	44.9 ± 20.5	45.0 ± 20.0	44.6 ± 21.6	.80
Dense scar (g)	24.1 ± 15.8	24.0 ± 15.3	24.2 ± 16.9	.87
Nondense scar (g)	20.8 ± 9.5	21.0 ± 9.6	20.3 ± 9.5	.41
Total scar (%LV)	40.2 ± 13.9	40.1 ± 13.7	40.3 ± 14.3	.84
Dense scar (%LV)	21.6 ± 12.9	21.6 ± 12.7	21.8 ± 13.4	.81
Nondense scar (%LV)	18.5 ± 0.1	18.6 ± 0.1	18.5 ± 0.1	.97

(Table 1 *continues*)

cal ethics committees, and all patients gave written informed consent before study participation.

The primary end point in the DERIVATE registry was defined as all-cause mortality. Secondary end points were cardiovascular death, SCD, aborted SCD (ie, appropriate ICD shock or antitachycardia pacing), and sustained ventricular tachycardia lasting at least 30 seconds or causing hemodynamic collapse in less than 30 seconds. The main end point of this retrospective analysis was a major adverse cardiac event (MACE), which was defined as the combination of SCD, aborted SCD, and sustained ventricular tachycardia.

The scar quantification machine algorithm can be used upon request. A prior article used the entire DERIVATE cohort (*n* = 2449), thereby including the 761 participants of the current study. This prior analysis dealt with the predictive power of right ventricular functional parameters for all-cause

Table 1 *(continued)*: Demographics, Medical History, and Clinical Cardiovascular Characteristics of Participants

Characteristic	All Patients (n = 761)	Learning Group (n = 515)	Testing Group (n = 246)	P Value*
Machine-determined imaging parameter				
Total scar (g)†	45.0 ± 19.0	45.9 ± 19.6	43.4 ± 18.0	.01
Dense scar (g)	23.2 ± 15.4	23.7 ± 15.6	22.2 ± 14.8	.28
Nondense scar (g)	21.9 ± 9.5	22.2 ± 9.9	21.1 ± 8.6	.16
Total scar (%LV)	37.8 ± 13.3	38.0 ± 13.5	37.4 ± 12.9	.56
Dense scar (%LV)	19.5 ± 12.2	19.7 ± 12.3	19.1 ± 12.0	.68
Nondense scar (%LV)	18.2 ± 6.5	18.2 ± 6.7	18.3 ± 6.3	.91
Quality LGE CMR				
Good	328 (43)	211 (41)	116 (47)	.08
Acceptable†	263 (35)	191 (37)	73 (30)	.04
Sufficient	75 (10)	49 (10)	26 (10)	.72
Borderline	50 (7)	34 (7)	16 (7)	>.99
Poor	45 (6)	30 (6)	15 (6)	.93

Note.—Data are numbers of patients, with percentages in parentheses. For continuous data, means ± SDs are presented. ACE/AT1 = angiotensin-converting enzyme/angiotensin receptor, BMI = body mass index, BSA = body surface area, CMR = cardiac MRI, ICD = implantable cardioverter defibrillator, LGE = late gadolinium enhancement, LVEF = left ventricular ejection fraction, LVEF-echo = left ventricular ejection fraction with echocardiography, MACE = major adverse cardiac event, NYHA = New York Heart Association, %LV = percentage of left ventricle, SCD = sudden cardiac death, VT = ventricular tachycardia.

* P values were determined by comparing the learning versus testing groups with use of Student independent t test.

† Denotes a significant $P < .05$.

‡ Defined as appropriate ICD intervention.

mortality and a composite of heart failure hospitalization and/or all-cause mortality (18).

CMR Examinations and Human Image Evaluation

All studies were acquired with 1.5-T machines (General Electric, Philips, Siemens Healthineers). CMR parameters and contrast media are presented in Table S2. Table S3 and Figure S1 describe how LGE image quality was graded. With the short-axis stack of LGE images, ground truth was determined in the core laboratory by two authors (F.G. and J.S., with 1 year and over 20 years of experience, respectively) for total left ventricle mass, dense scar (scar signal: >5 SDs above the reference mean of remote normal myocardium), total scar (scar signal: >2 SDs above the reference mean), and nondense scar (2 SDs ≤ scar signal ≤ 5 SDs of the reference mean) with use of GTvolume software (GyroTools, version 2.2.7). In difficult cases (typically with poor-sufficient LGE image quality), ground truth was generated in consensus (by F.G. and J.S.). The ground truth results were then used for the machine comparison (see also Appendix S1). Intra- and interobserver variabilities (F.G. vs F.G. and F.G. vs A.G.P., with over 5 years of experience) were also assessed. All observers were blinded to the demographics and outcome data. The thresholds of over 5 SDs and over 2 SDs were chosen as they were already successfully applied by others (5,7,19).

Machine Learning Algorithm

For our segmentation task, we implemented a fully convolutional neural network (20)—specifically, a Ternaus network (21,22) (Fig 2). For details of the network, data preprocessing, and network training, see Appendix S1.

After ground truth production, the study sample was randomly divided into a learning group (n = 515) and a testing group (n = 246). The network was trained with the LGE images of the learning group, with the other third kept separate (off site) until the final evaluation. During training, the learning group was split into training (n = 413) and validation (n = 102) sets. The network was trained only on the training portion of the learning group, and its performance was evaluated on the validation portion to guide design decision (see also Appendix S1). Only when the final network was trained did we evaluate the network performance with use of the testing group.

To improve the final predictions of the trained network, we made use of a test-time augmentation (TTA) approach. For details of the TTA approach and the four prediction outputs, see Appendix S1. As the TTA output performed best (Fig S2), only these outputs were used for further outcome prediction analyses.

Statistical Analyses

Regarding sample size calculations, we want to refer the reader to Guaricci et al (10), which defined one variable per 10 events to include, thus, allowing the inclusion of a maximum of three variables in the multivariable Cox proportional hazard models (see later section). Patient demographic data exhibited a normal distribution as determined by the Kolmogorov-Smirnov test. The Student independent t, χ^2, or Fisher exact tests were used as appropriate to compare variables between groups. For comparisons of human intra- and interobserver performances as well as the performance of human versus machine analysis, the median absolute error, IQR, and outliers were calculated. Dice metrics were also

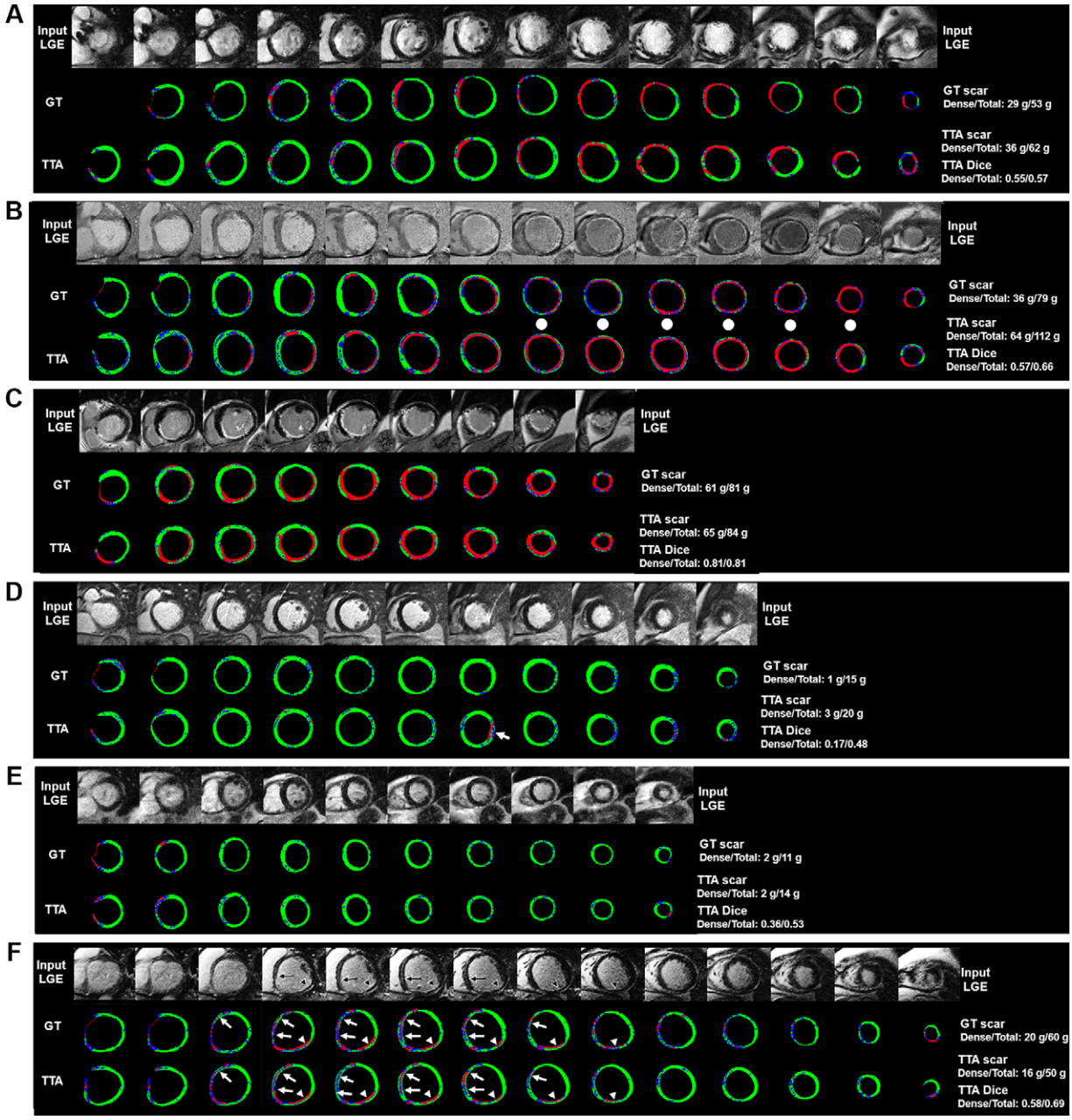

Figure 3: Examples of short-axis late gadolinium enhancement (LGE) cardiac MRI scans with scar mass quantification by humans (ground truth [GT]) and the machine algorithm (test-time augmented [TTA]). For all images, green represents normal remote myocardium, red represents total-2SD-scar, and blue represents dense-5SD-scar. TTA dice represents Dice similarity coefficients, where 0 means no overlapping of the machine-generated contours with the ground truth contours and 1 means 100% overlap. **(A)** Images in a 68-year-old man show a left ventricular ejection fraction (LVEF) with echocardiography of 37% and a machine-quantified (TTA) dense scar greater than 20.3 g (36 g). This participant had no implantable cardioverter defibrillator (ICD), and sudden cardiac death (SCD) occurred 1 month after cardiac MRI. **(B)** Images in a 57-year-old man show a LVEF with echocardiography of 20% and machine-quantified dense scar mass of 64 g. SCD occurred 2.2 years after cardiac MRI. If no normal reference myocardium was present in a section (white points), the reference myocardium of the next available section was used for threshold calculations. The lack of reference myocardium in several sections may explain the difference in scar mass as determined by human versus machine in this case. **(C)** Images in a 55-year-old man show a LVEF with echocardiography of 35% and a machine-determined dense scar mass of 65 g. Participant had sustained ventricular tachycardia and an appropriate ICD shock 11 months after cardiac MRI. **(D)** Images in a 64-year-old man show a LVEF withby echocardiography of 31% and a machine-determined dense scar mass of 3 g. No major adverse cardiac event (MACE) was reported. ICD implantation and nonsustained ventricular tachycardia occurred 3.7 years after cardiac MRI. Image includes a ghosting artifact (arrow), which was "corrected" by the human observer in the ground truth analysis but falsely identified as a scar by the machine algorithm. **(E)** Images in a 78-year-old man show a LVEF with echocardiography of 47% and a machine-determined dense scar mass of 2 g. No MACE was reported over 1.6 years after cardiac MRI. **(F)** Images in a 67-year-old man show false-negative machine algorithm assessment. Both human and machine scar quantification yield low scar burden below thresholds for MACE (20 g and 16 g, respectively) in this patient with a low LVEF with echocardiography of 15%. The participant suffers from coronary artery disease with a small subendocardial inferior scar (arrowheads). The LGE images also show a midline sign (arrows), a feature known to be associated with MACE (35), that was detected by both human and machine analysis, and which may represent an arrhythmic substrate. Participant had ICD implantation at 0.5 years and MACE (sustained ventricular tachycardia) at 1.7 years after cardiac MRI.

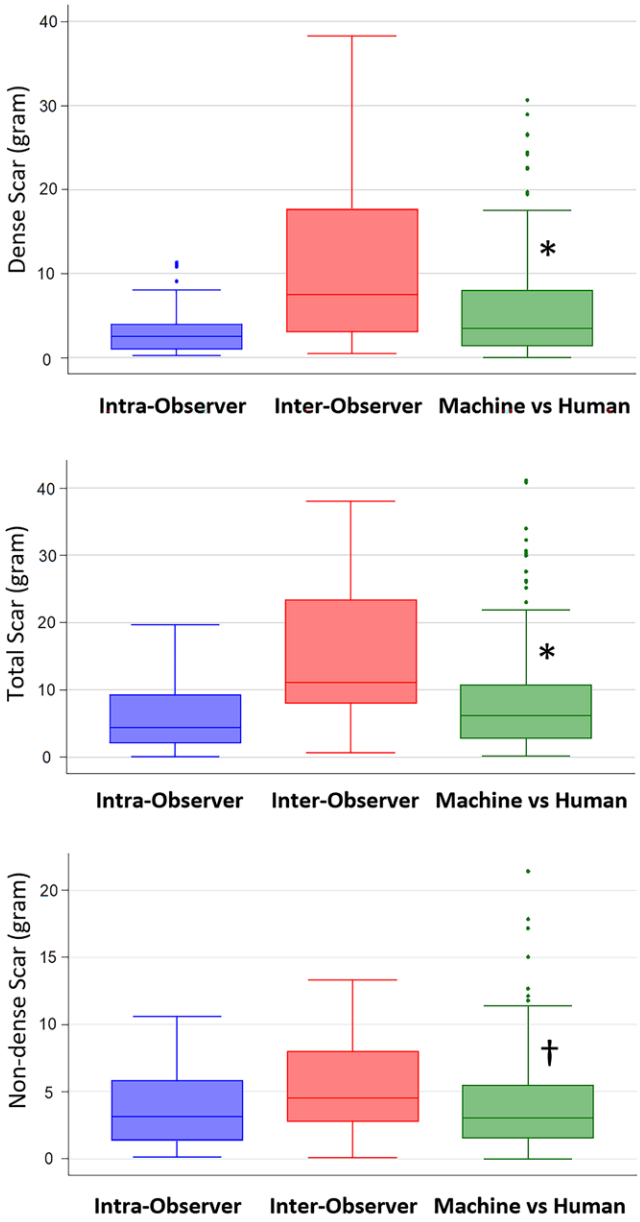

Figure 4: Box and whisker plots show median absolute error of dense scar, total scar, and nondense scar mass as determined by humans (intraobserver in blue [n = 40] and interobserver in red [n = 40]) as well as for the human versus machine comparison (green [n = 246]). The median absolute errors of the novel machine algorithm are not different versus errors of the intraobserver comparison, with medians of 3.54 versus 2.57 (P = .08), 6.41 versus 4.37 (P = .09), and 3.23 versus 3.62 (P = .63) for dense, total, and nondense scar (gram), respectively, but are lower compared with the interobserver errors, with medians of 3.54 versus 7.54, 6.41 versus 11.14, and 3.23 versus 4.56 for dense, total, and nondense scar, respectively. * P < .001. † P < .02 (Mann-Whitney test). Whiskers on box plots represent the 1.5 quartile boundaries, and the horizontal line in the box plot represents the median. Circles outside the box plots represent outliers. If the intraobserver and interobserver errors were compared with errors of the machine algorithm of the same 40 patients (data not shown), no differences were found versus intraobserver (P = .73, P = .66, and P = .18 for dense, total, and nondense scar, respectively) and errors were significantly lower versus interobserver (P < .001, P < .001, and P < .005, respectively) (Mann-Whitney test).

calculated. Comparisons between groups were performed by means of the nonparametric Mann-Whitney test.

End of follow-up or censoring event-free survival was estimated with use of the Kaplan–Meier method, and survival curves were compared with use of the log-rank test.

Univariable and multivariable Cox proportional hazard models were used to assess the association of demographic and imaging data with the risk of MACE. Different multivariable models were adapted to evaluate whether human-based or machine-based scar measures improve the currently established guideline criterion (LVEF of ≤35% and dyspnea New York Heart Association II or III, model 0). Model 1 included guideline criterion and human-determined dense scar quantification, model 2 included guideline criterion and machine-determined dense scar quantification, and model 3 included guideline criterion and human- and machine-determined dense scar quantification. Models 1a, 2a, and 3a include the same groups but use total scar quantification. To assess the performance of the multivariable models, estimated areas under the receiver operating characteristic curve (AUCs) for the time-dependent receiver operator characteristic curves were derived. Univariable and multivariable Cox models were also calculated considering noncardiovascular death as competing risk. For details, see Appendix S2.

Two-sided P < .05 was considered significant for statistical tests, and Bonferroni correction was used when appropriate. All the analyses were performed by an author (V.L., with over 10 years of experience) with use of STATA version 14 and R version 3.6.2 (R Core Team, 2019) (23).

Results

Patient Characteristics and LGE Data Quality

The study population consisted of 761 patients (671 men) with a mean age ± SD of 65 years ± 11. Further demographics are given in Table 1. Over a mean follow-up of 2.9 years ± 1.9, MACE occurred in 83 of the 761 participants, 56 from the learning group and 27 from the testing group (P = .53). Of the patients formally fulfilling the guidelines criteria for primary ICD implantation (11–13), 67% were treated by ICD implantation in both the learning and testing group.

Inadequate LGE quality led to the exclusion of five participants from statistical analysis (Fig 1). In the total population of 761 patients, good or acceptable quality was found in 591 patients (77.7%) (Table 1). Distribution of LGE quality grades was similar in the participants stratified by learning and testing groups; only one difference was found for acceptable quality (191 artbitrary units [37%] vs 73 arbitrary units [30%], respectively; P = .04) (Table 1). For interobserver reproducibility of quality grading, see Table S4.

Comparison of Scar Quantification by Humans and by the TTA Machine Algorithm

The TTA output demonstrated the best performance compared with humans (Fig S2). Representative examples of LGE images with human results and TTA machine outputs are given in Figure 3.

Table 2: Univariable Cox Analysis for Participant Demographics and Clinical Cardiac Characteristics Associated with Major Adverse Cardiac Event

Characteristic	HR	*P* Value	95% CI
Demographics			
Age (y)	1.02	.16	0.99, 1.05
Sex (M)	2.14	.06	0.96, 4.81
BMI (kg/m^2)	0.97	.67	0.85, 1.11
BSA (m^2)	1.14	.87	0.23, 5.72
Family history	1.33	.33	0.75, 2.35
Smoking	0.59	.09	0.32, 1.08
Hyperlipidemia	0.60	.07	0.35, 1.04
Diabetes	1.21	.53	0.67, 2.22
Hypertension	0.95	.86	0.55, 1.64
NYHA class*	1.39	.02	1.05, 1.85
Left bundle branch block	0.78	.47	0.40, 1.53
Echocardiography			
LVEDVI (mL/m^2)*	1.01	.001	1.01, 1.02
LVESVI (mL/m^2)*	1.02	<.001	1.01, 1.03
LVEF-echo (%)*	0.94	<.001	0.91, 0.97
LVEF-echo (threshold ≤35%)*	4.68	<.001	2.19, 9.98
LVEF-echo (≤35%) and NYHA class II or III*†	3.99	<.001	2.24, 7.13
CMR function and volume			
LVEDVI (mL/m^2)*	1.01	.005	1.00, 1.01
LVESVI (mL/m^2)*	1.01	.003	1.00, 1.01
LVSVI (mL/m^2)	0.99	.12	0.98, 1.00
LVEF (%)*	0.95	<.001	0.92, 0.97
RVEDVI (mL/m^2)*	1.01	.04	1.00, 1.03
RVESVI (mL/m^2)*	1.02	.002	1.01, 1.03
RVSVI (mL/m^2)	0.99	.20	0.98, 1.01
RVEF (%)*	0.97	.003	0.95, 0.99
LGE CMR quality			
Acceptable versus good	0.62	.15	0.32, 1.19
Sufficient versus good	0.65	.41	0.23, 1.19
Borderline versus good	0.61	.50	0.15, 2.54
Poor versus good	0.91	.90	0.22, 3.82
CMR scar, human (GT)			
LVmass (g)	1.01	.12	1.00, 1.02
totalScar%$_{hum}$ (%LV)*	9.20	.03	1.22, 69.57
denseScar%$_{hum}$ (%LV)*	9.58	.03	1.24, 73.78
nondenseScar%$_{hum}$ (%LV)	1.02	.25	0.99, 1.05
totalScar$_{hum}$ (g)*	1.02	.01	1.00, 1.03
denseScar$_{hum}$ (g)*	1.02	.01	1.00, 1.03
nondenseScar$_{hum}$ (g)	1.87	.80	0.02, 266.99
CMR scar, machine (TTA)			
LVmass$_{mach}$ (g)*	1.01	.04	1.00, 1.02
totalScar%$_{mach}$ (%LV)*	12.57	.02	1.65, 95.90
denseScar%$_{mach}$ (%LV)*	17.78	.008	2.09, 151.16
nondenseScar%$_{mach}$ (%LV)	1.02	.14	0.99, 1.06
totalScar$_{mach}$ (g)*	1.02	.002	1.01, 1.04
denseScar$_{mach}$ (g)*	1.06	.002	1.01, 1.04
nondenseScar$_{mach}$ (g)	1.22	.93	0.01, 130.49
Threshold (machine)			
Total scar (>30.7 g)*‡	3.61	.002	1.63, 8.02
Dense scar (>20.3 g)*‡	2.35	.003	1.33, 4.15

(Table 2 *continues*)

Table 2 *(continued)*: Univariable Cox Analysis for Participant Demographics and Clinical Cardiac Characteristics Associated with Major Adverse Cardiac Event

Note.—Table shows results for testing group two-dimensional late gadolinium enhancement (LGE) images (n = 216). BMI = body mass index, BSA = body surface area, CMR = cardiac MRI, GT = ground truth, HR = hazard ratio, LV = left ventricle, LVEDVI = left ventricular end-diastolic volume index, LVEF = left ventricular ejection fraction, LVEF-echo = left ventricular ejection fraction with echocardiography, LVESVI = left ventricular end-systolic volume index, LVSVI = left ventricular stroke volume index, NYHA = New York Heart Association, %LV = percentage of left ventricle, RVEDVI = right ventricular end-diastolic volume index, RVEF = right ventricular ejection fraction, RVESVI = right ventricular end-systolic volume index, RVSVI = right ventricular stroke volume index, TTA = test-time augmentation.

* Denotes a significant $P < .05$.

† Criterion for implantable cardioverter defibrillator implantation, according to international guidelines (11–13).

‡ Data are areas under the receiver operator characteristic curve.

With use of the testing data set (n = 246), the comparison of machine versus core laboratory-determined human performance of scar quantification yielded median absolute error of 2.9 g (IQR, 1.1–5.5 g) for dense scar and 4.2 g (IQR, 1.8–7.3 g) for total scar (Fig 4, Table S5). For the interobserver (Fig 4) and intraobserver (Fig 4) analyses, the median absolute error is 11.1 g (IQR, 8.1–22.8 g) and 4.4 g (IQR, 2.2–8.5 g) for total scar, respectively, and 7.5 g (IQR, 3.0–17.1 g) and 2.5 g (IQR, 1.0–4.1 g) for dense scar, respectively. As shown in Figure 4, the quantification performance of the developed machine algorithm is comparable to the intraobserver performance (difference of median absolute errors for intraobserver vs machine: P = .08, P = .09, P = .63, for dense, total, and nondense scar, respectively) and is superior to the interobserver performance with significantly lower median absolute errors (difference of median absolute errors for interobserver vs machine: $P < .001$, $P < .001$, $P < .02$, for dense, total, and nondense scar, respectively). For Bland-Altman analyses and scar quantification performance of the machine algorithm in the learning group, see Figures S3–S6.

In comparing the machine algorithm versus humans, the Dice similarity coefficients were 80.0% ± 5.5 for normal myocardium, 62.0% ± 11.4 for total scar, 56.8% ± 17.2 for dense scar, and 64.7% ± 12.0 for nondense scar (analysis of variance, $P < .001$).

Subgroup analyses demonstrated similar performance of the machine algorithm versus humans independent of LGE image quality, as median absolute errors between the machine algorithm versus humans were similar in the group of participants with good quality LGE images (best grading [n = 116]) versus the composite group of acceptable, sufficient, borderline, and poor quality (ie, "non-good" quality [n = 130]) for total, dense, and nondense scar (Fig S7A) (P = .051, P = .13, P = .052, respectively) (Mann-Whitney test). Similarly, no differences were found for machine performance versus humans based on

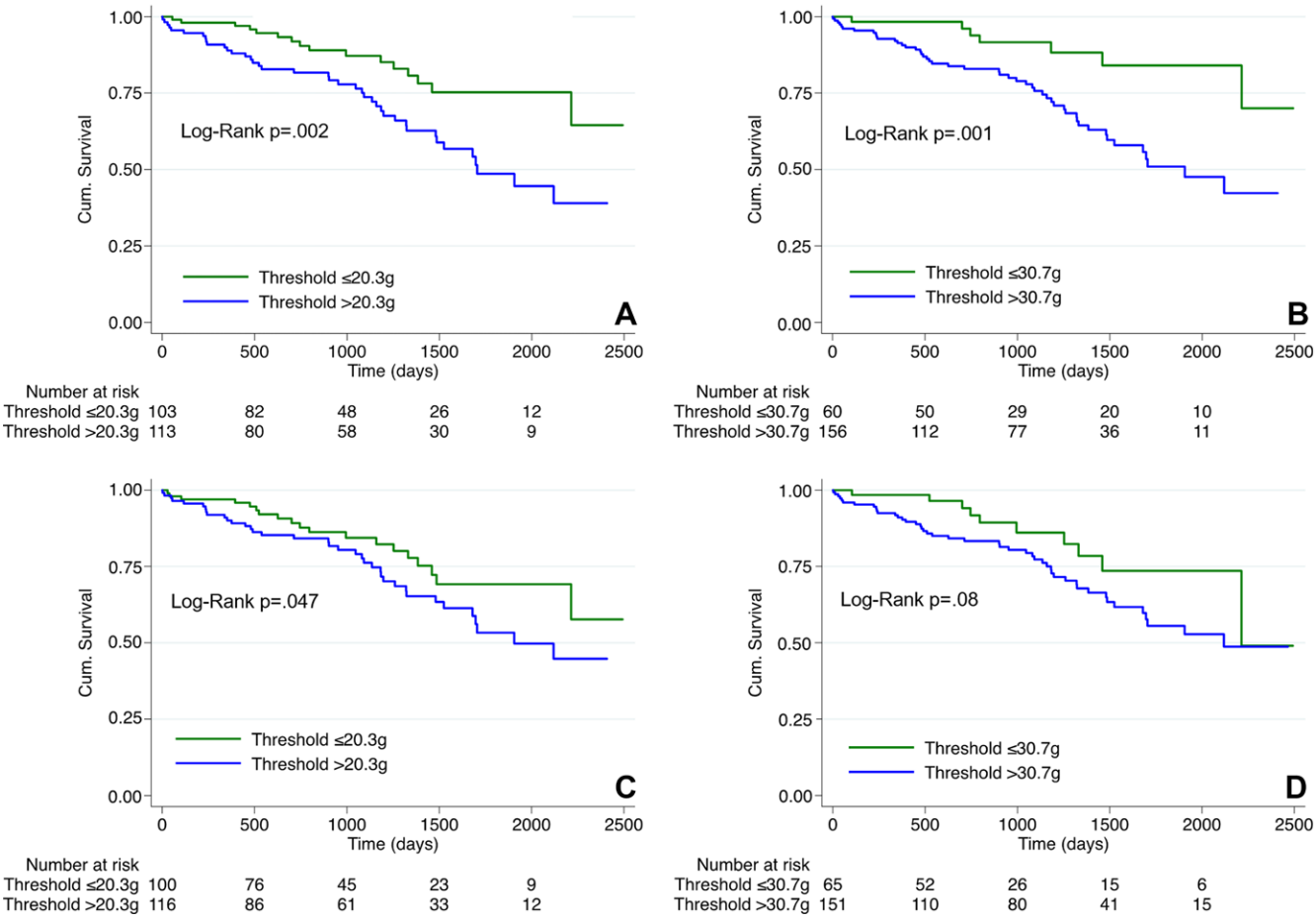

Figure 5: Graphs show major adverse cardiac event (MACE) risk prediction by absolute late gadolinium enhancement scar mass in human versus machine analysis. Kaplan–Meier curves show the MACE prediction for the thresholds of machine-determined **(A)** dense scar and **(B)** total scar mass and human-determined **(C)** dense scar and **(D)** total scar mass. The log-rank *P* values were *P* = .002, *P* = .001, *P* = .047, and *P* = .08, for **A**, **B**, **C**, and **D**, respectively. Cum. = cumulative.

participant sex (Fig S7B) (*P* = .15, *P* = .21, *P* = .09, for total, dense, and nondense scar, respectively) (Mann-Whitney test) or age (Fig S7C) (*P* = .67, *P* = .18, *P* = .69, for total, dense, and nondense scar, respectively) (Mann-Whitney test).

In the testing group (*n* = 246) (Fig 1), the three-dimensional LGE images were inferior in quality versus two-dimensional LGE images (good or acceptable quality in 61% vs 80%, respectively, *P* < .001). Therefore, for MACE prediction modeling, the two-dimensional LGE data were used (final testing group [*n* = 216]) (Fig 1).

MACE Prediction Based on Demographic, Echocardiographic, and CMR Parameters

With use of univariable Cox analysis, New York Heart Association class showed a significant association with the risk of MACE (hazard ratio, 1.39; 95% CI: 1.05, 1.85; *P* = .02) (Table 2). The guideline criterion correlated significantly with the risk of MACE (hazard ratio, 3.99; 95% CI: 2.24, 7.13; *P* < .001) as did most echocardiographic and CMR-derived parameters (LVEF echo: hazard ratio = 0.94; 95% CI: 0.91, 0.97, *P* < .001; LVEF CMR: hazard ratio = 0.95, 95% CI: 0.92, 0.97, *P* < .001) (Table 2). In addition, all CMR scar parameters analyzed by humans or by the machine algorithm were found to be predictors of MACE except nondense scar mass (human: hazard ratio = 1.87, 95% CI: 0.02, 266.99, *P* = .80; machine: hazard ratio = 1.22, 95% CI: 0.01, 130.49, *P* = .93) (Table 2). Univariable Cox analyses with competitive risk assessment are given in Table S6.

Event-free survival curves for the thresholds of total scar and dense scar as determined by humans and the machine algorithm showed that the machine-based results are associated with event-free survival for dense scar (*P* = .002) and total scar (*P* = .001) (Fig 5). Human analysis was also associated with the threshold of dense scar (*P* = .047) (Fig 5).

Age, LVEF (echocardiography and CMR), and the thresholds for total scar determined by both humans and the machine algorithm were predictive of all-cause mortality, the primary end point of the DERIVATE study (Table S7).

Dense scar threshold determined by machine (model 2) or total scar threshold determined by machine (model 2b) conferred incremental value over the guideline criterion (model 0) for the association with MACE (AUC, 0.63 vs 0.68; *P* = .02 and AUC, 0.63 vs 0.67; *P* = .01, respectively) (Table 3). This superiority was not observed for dense scar threshold determined by human (model 1) nor for total scar threshold determined by human (model 1b) versus the guideline criterion (model 0 with AUC, 0.63 vs 0.65, *P* = .36 and AUC, 0.63 vs

Table 3: Multivariable Cox Model for Cardiac Characteristics Associated with Major Adverse Cardiac Event

Characteristic	HR	*P* Value	95% CI	AUC $_{\text{Time-dependent ROC}}$ *
Multivariable Cox models–dense scar				
Model 0				
Guideline criterion†	3.99	<.001	2.24, 7.13	0.63 (0.52, 0.74)
Model M1–dense scar, human-determined				0.65 (0.54, 0.75) [*P* = .36 vs M0]
Guideline criterion†	3.63	<.001	2.076, 6.34	...
Threshold over 20.3 g	1.26	.42	0.72, 2.22	…
Model M2–dense scar, machine-determined				0.68 (0.58, 0.78) [*P* = .02 vs M0] [*P* = .13 vs M1]
Guideline criterion†	3.39	<.001	1.95, 5.91	...
Threshold over 20.3 g	1.79	.050	1.00, 3.19	…
Model M3–dense scar, human- and machine-determined				0.69 (0.59, 0.78) [*P* = .02 vs M0] [*P* = .18 vs M1]
Guideline criterion†	3.49	<.001	2.00, 6.11	...
Threshold over 20.3 g, human	0.60	.24	0.26, 1.41	…
Threshold over 20.3 g, machine	2.64	.03	1.10, 6.37	…
Multivariable Cox models–total scar				
Model 0				
Guideline criterion†	3.99	<.001	2.24, 7.13	0.63 (0.52, 0.74)
Model M1b–total scar, human-determined				0.65 (0.54, 0.75) [*P* = .09 vs M0]
Guideline criterion†	3.65	<.001	2.10, 6.34	...
Threshold–totalScar$_{\text{hum}}$ (>30.7 g)	1.37	.38	0.681, 2.746	…
Model M2b–total scar, machine-determined				0.67 (0.58, 0.76) [*P* = .01 vs M0] [*P* = .36 vs M1b]
Guideline criterion†	3.10	<.001	1.76, 5.47	...
Threshold–totalScar$_{\text{mach}}$ (>30.7 g)	2.31	.047	1.01, 5.29	…
Model M3b–total scar, human- and machine-determined				0.67 (0.58, 0.76) [*P* = .03 vs M0] [*P* = .64 vs M1b]
Guideline criterion†	3.08	<.001	1.75, 5.44	...
Threshold over 30.7 g, human	0.79	.58	0.34, 1.83	…
Threshold over 30.7 g, machine	2.70	.05	0.99, 7.38	…

Note.—Table shows results for testing group two-dimensional late gadolinium enhancement images (*n* = 216). AUC = area under the receiver operating characteristic curve, HR = hazard ratio, ROC = receiver operator characteristic curve.

* Data in parentheses are 95% CIs, and data in brackets are *P* values.

† Left ventricular ejection fraction with echocardiography (≤35%) and New York Heart Association class II or III were criteria for implantable cardioverter defibrillator implantation, according to international guidelines (11–13).

0.65, *P* = .09, respectively) (Table 3). For illustration of the model 2 performance, the Kaplan–Meier curves in Figure 6 show that the combination of the guideline criterion with dense scar determined by machine discriminates different risk groups (log-rank test, *P* < .001). Model 3 included both the human- and machine-determined dense scar threshold and performed better than the guideline criterion alone (model 0) with AUC of model 0 of 0.63 versus AUC of model 3 of 0.69, *P* = .02 (Table 3). This model 3 did not perform better than model 1 (guideline criterion and dense scar threshold determined by human) (AUC of 0.69 for model 3 vs AUC of 0.65 for model 1, *P* = .18). The same lack of superiority was observed for total scar mass (model 3b) versus model 1b (AUC, 0.67 vs 0.65; *P* = .64). Modeling with consideration of competing risks confirmed these results (Table S8).

Discussion

The goals of the study were to test whether postinfarct scar with late gadolinium enhancement (LGE) cardiac MRI (CMR) can be quantified fully automatically by machines and compare the ability of LGE CMR scar analyzed by humans and machines to predict arrhythmic events. To this end, a machine learning–based algorithm was developed that quantifies myocardial scars without any user interaction, and no endo- and epicardial contours are needed. The algorithm yields a small median absolute error versus the ground truth of 2.9 g (IQR, 1.1–5.3 g) for dense scar and 4.2 g (IQR, 1.8–7.3 g) for total scar.

The machine-based results of dense and total scar mass in grams were both associated with the risk of MACE (hazard ratio, 1.03 and 1.02, respectively; *P* = .002 for both). Kaplan–Meier curves demonstrate the ability to stratify survival for the thresholds of 20.3 g and 30.7 g for machine-determined dense scar and total scar, respectively (*P* < .001) (log-rank test). The human analysis yielded scar quantification that was also associated with the risk of MACE (hazard ratio, 1.02; *P* = .01 for both, total and dense scar). Interestingly, in multivariable Cox regression analyses (models 0–3), machine-based thresholds of dense scar and total scar (models 2 and 2b) improved the guideline criterion–based model (AUC, 0.68 and 0.67 vs 0.63; *P* = .02 and *P* = .01, respectively), while the human-based

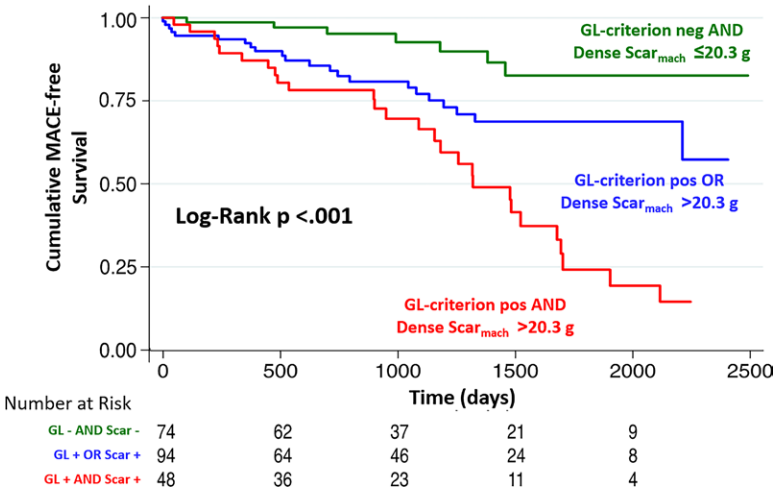

Figure 6: Graph shows major adverse cardiac event (MACE) risk prediction with use of the combination of the guideline (GL) criterion– and machine-based dense scar mass of late gadolinium enhancement cardiac MRI. Kaplan–Meier curves for MACE prediction show that the combination of these parameters discriminates between different risk groups (log-rank test, $P < .001$). The positive (pos) guideline criterion is defined as left ventricular ejection fraction of 35% or smaller and dyspnea New York Heart Association class II or III. The green survival curve represents a combination of negative (neg) guideline criterion and machine-determined dense scar mass (Dense Scar$_{mach}$) of less than or equal to 20.3 g; the blue survival curve represents a combination of positive guideline criterion or machine-determined dense scar mass of more than 20.3 g; the red survival curve represents a combination of positive guideline criterion and machine-determined dense scar mass of more than 20.3 g.

scar results did not (AUC, 0.65 and 0.65 vs 0.63; $P = .36$ and $P = .09$, respectively).

To our knowledge, current LGE analyses generally used some user input, such as human tracings of endo- and epicardial contours (24,25) or contours of cine images (26). Furthermore, previous studies with use of LGE analysis did not quantify scars (27), only used three-dimensional data (28), and did not assess the predictive performance of the machine learning–based algorithm (24–28). The novel TTA machine algorithm presented here demonstrates similar performance in comparison to human analyses irrespective of LGE image quality for total, dense, and nondense scar ($P = .051$, $P = .13$, $P = .052$, respectively) (Mann-Whitney test), gender ($P = .15$, $P = .21$, $P = .09$, respectively), or age ($P = .67$, $P = .18$, $P = .69$, respectively) of participants, and it eliminates any scar analysis variability even in multicenter data. In addition, the algorithm reduces the tedious LGE analysis time down to 0.5 seconds per heart.

Several single-center studies demonstrated the predictive value of human-determined scar for MACE (7,19,29–31), and the current study extends this knowledge to a multicenter, multivendor setting. Regarding the fully automatic scar analysis by a machine learning algorithm, this is, to our knowledge, the first time that these measures are shown to confer incremental value over the guideline criterion (the combination of echocardiographic LVEF of 35% or smaller and New York Heart Association class II or III [11–13]) for the prediction of MACE. The current findings may suggest the possibility to hand over time-consuming scar analyses from humans to machines.

Applying machine learning methods to directly predict MACE (instead of learning to segment scars) might perform better than the presented approach. Such a strategy was applied in a three-center learning cohort, and the AUC to predict arrhythmic SCD was 0.72 (32), thus, slightly higher than the best AUC for MACE prediction of 0.69 found in our study. Ventricular tachycardias in postmyocardial infarction scar are based on reentry circuits through conducting channels in the scar tissue. Nondense scar typically is the substrate for such channels, while dense scar forms the borders. Thus, both scar types are relevant to provide the arrhythmic substrate (33). Several studies support nondense scar as most predictive for MACE (7,29,31,34), while others (19) and the current study favor dense scar. Future studies are warranted to address this question in detail.

Our study had several limitations. First, as in any registry, referral biases cannot be excluded (eg, decompensated patients with severe dyspnea [and difficulties to perform breath holds] might be underrepresented). Second, our algorithm was tested in a held-out population of the DERIVATE cohort, and no external population was used for testing. Thus, the findings of this study should only be applied to populations that fulfill the inclusion criteria of this trial (including the LGE CMR parameters) (10). The three-dimensional LGE images were inferior in quality versus two-dimensional LGE images (good or acceptable quality in 61% vs 80%, $P < .001$) and were, therefore, not used for outcome prediction modeling, and consequently, the prediction performance is restricted to two-dimensional LGE images. Third, at the time of study inclusion, novel T1-mapping techniques or sophisticated three-dimensional LGE techniques that were acquired during free breathing were not available in all centers, which could have improved prediction performance and interobserver reproducibility, respectively. We would like to note that it was the aim of the study to investigate the prognostic power of scar quantification with breath-hold LGE CMR that could be readily applied in general cardiology routine. Fourth, based on the current data, we cannot investigate potential mechanisms to explain the trend of better outcome prediction of the machine algorithm versus humans. Fifth, event documentation was accomplished by the local physicians, and no independent adjudication committee was established, which is certainly a limitation. Finally, our study was not powered to assess the influence of different machine vendors, contrast medium doses or type, nor pulse sequence types on machine learning–based scar quantification or outcome prediction.

In this large international, multicenter, multivendor setting, scar with late gadolinium enhancement cardiac MRI was fully, automatically analyzed by a machine learning–based algorithm. This approach improved the current prediction model that uses guideline-based risk criteria to identify candidates for implantable cardioverter defibrillators. Future prospective trials are warranted to evaluate the performance and added value of machine learning–based scar quantification for better major adverse cardiac event prediction and improved arrhythmia management of patients with coronary artery disease.

Author contributions: Guarantors of integrity of entire study, **F.G., L.F., S. Carigi, G. Casavecchia, S. Censi, M.G., M.L., A.M., A.N., A.V., G.P., J.S.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **F.G., T.J., A.I.G., A.G.P., L.F., M.G.R., G.D.A., R.A., J.B., G. Camastra, S. Carigi, G.D.G., M.F., M.G., C.L., M.L., A.M., G.M., A.N., S.P., A.V., G.P., S.K., J.S.**; clinical studies, **A.I.G., A.G.P., L.F., D.A., M.G.R., G.D.A., J.B., G. Camastra, S. Carigi, N.C., G. Casavecchia, S. Censi, M.D.R., M.F., N.G., A.G., M.G., M.L., J.L.T., A.M., C.M., A.N., S.P., S.B., U.J.S., A.V., P.G.M., G.P., J.S.**; experimental studies, **T.J., A.I.G., A.G.P., G.D.A., S. Carigi, G. Cicala, M.G., A.N., A.V., G. Crelier, P.G.M., G.P.**; statistical analysis, **F.G., T.J., V.L., A.G.P., G.D.A., J.B., S. Carigi, G. Casavecchia, M.D.L., M.G., A.N., A.V., G.P., S.K.**; and manuscript editing, **F.G., T.J., V.L., A.I.G., A.G.P., D.A., M.G.R., G.D.A., J.B., S. Carigi, N.C., C.N.D.C., M.D.L., M.F., N.G., M.G., M.L., G.M., A.N., S.P., U.J.S., A.V., P.G.M., G.P., S.K., J.S.**

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: F.G. No relevant relationships. **T.J.** Financial support of the PHRT SWISSHEART Failure Network through institution (ETH Zurich). **V.L.** No relevant relationships. **A.I.G.** No relevant relationships. **A.G.P.** No relevant relationships. **L.F.** No relevant relationships. **D.A.** No relevant relationships. **M.G.R.** No relevant relationships. **G.D.A.** No relevant relationships. **R.A.** No relevant relationships. **J.B.** No relevant relationships. **G. Camastra** No relevant relationships. **S. Carigi** No relevant relationships. **N.C.** No relevant relationships. **G. Casavecchia** No relevant relationships. **S. Censi** No relevant relationships. **G. Cicala** No relevant relationships. **C.N.D.C.** Research grant from Siemens; consulting fees from Covanos. **M.D.L.** No relevant relationships. **G.D.G.** No relevant relationships. **M.D.R.** No relevant relationships. **M.F.** No relevant relationships. **N.G.** No relevant relationships. **A.G.** No relevant relationships. **M.G.** No relevant relationships. **C.L.** No relevant relationships. **M.L.** No relevant relationships. **J.L.T.** No relevant relationships. **A.M.** No relevant relationships. **C.M.** No relevant relationships. **G.M.** Support for attending meetings and/or travel from Bracco; chairperson for Communication & New Media Committee of ESCR. **A.N.** No relevant relationships. **S.P.** No relevant relationships. **S.S.** No relevant relationships. **U.J.S.** Institutional research support and/or honoraria for speaking and consulting from Astellas, Bayer, Bracco, Elucid BioImaging, General Electric, Guerbet, Heartflow, and Siemens Healthineers. **A.V.** No relevant relationships. **G. Crelier** Employed with GyroTools. **P.G.M.** Consultancy fees from Perspectum Diagnostics. **G.P.** No relevant relationships. **S.K.** No relevant relationships. **J.S.** Supported by Bayer Healthcare Schweiz; member of the advocacy committee of the Society of Cardiovascular Magnetic Resonance.

References

1. Roth GA, Huffman MD, Moran AE, et al. Global and regional patterns in cardiovascular mortality from 1990 to 2013. Circulation 2015;132(17):1667–1678.
2. Mehta D, Curwin J, Gomes JA, Fuster V. Sudden death in coronary artery disease: acute ischemia versus myocardial substrate. Circulation 1997;96(9):3215–3223.
3. Kim RJ, Wu E, Rafael A, et al. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. N Engl J Med 2000;343(20):1445–1453.
4. Knuesel PR, Nanz D, Wyss C, et al. Characterization of dysfunctional myocardium by positron emission tomography and magnetic resonance: relation to functional outcome after revascularization. Circulation 2003;108(9):1095–1100.
5. Kwon DH, Asamoto L, Popovic ZB, et al. Infarct characterization and quantification by delayed enhancement cardiac magnetic resonance imaging is a powerful independent and incremental predictor of mortality in patients with advanced ischemic cardiomyopathy. Circ Cardiovasc Imaging 2014;7(5):796–804.
6. Scott PA, Morgan JM, Carroll N, et al. The extent of left ventricular scar quantified by late gadolinium enhancement MRI is associated with spontaneous ventricular arrhythmias in patients with coronary artery disease and implantable cardioverter-defibrillators. Circ Arrhythm Electrophysiol 2011;4(3):324–330.
7. Zegard A, Okafor O, de Bono J, et al. Myocardial Fibrosis as a Predictor of Sudden Death in Patients With Coronary Artery Disease. J Am Coll Cardiol 2021;77(1):29–41.
8. Disertori M, Rigoni M, Pace N, et al. Myocardial Fibrosis Assessment by LGE Is a Powerful Predictor of Ventricular Tachyarrhythmias in Ischemic and Nonischemic LV Dysfunction: A Meta-Analysis. JACC Cardiovasc Imaging 2016;9(9):1046–1055.
9. Scott PA, Rosengarten JA, Curzen NP, Morgan JM. Late gadolinium enhancement cardiac magnetic resonance imaging for the prediction of ventricular tachyarrhythmic events: a meta-analysis. Eur J Heart Fail 2013;15(9):1019–1027.
10. Guaricci AI, Masci PG, Lorenzoni V, Schwitter J, Pontone G. CarDiac MagnEtic Resonance for Primary Prevention Implantable CardioVerter DebrillAtor ThErapy international registry: Design and rationale of the DERIVATE study. Int J Cardiol 2018;261:223–227.
11. Priori SG, Blomström-Lundqvist C, Mazzanti A, et al. 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC). Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J 2015;36(41):2793–2867.
12. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021;42(36):3599–3726.
13. Al-Khatib SM, Stevenson WG, Ackerman MJ, et al. 2017 AHA/ACC/HRS Guideline for Management of Patients With Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Circulation 2018;138(13):e210–e271.
14. Stecker EC, Vickers C, Waltz J, et al. Population-based analysis of sudden cardiac death with and without left ventricular systolic dysfunction: two-year findings from the Oregon Sudden Unexpected Death Study. J Am Coll Cardiol 2006;47(6):1161–1166.
15. Jiang B, Guo N, Ge Y, Zhang L, Oudkerk M, Xie X. Development and application of artificial intelligence in cardiac imaging. Br J Radiol 2020;93(1113):20190812.
16. Bluemke DA. Radiology in 2018: Are You Working with AI or Being Replaced by AI? Radiology 2018;287(2):365–366.
17. Yancy CW, Jessup M, Bozkurt B, et al. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2013;62(16):e147–e239.
18. Al'Aref SJ, Malkawi A, Pontone G. Reply to: Letter to the Editor regarding 'Cardiac magnetic resonance for prophylactic implantable-cardioverter defibrillator therapy international study: prognostic value of cardiac magnetic resonance-derived right ventricular parameters substudy'. Eur Heart J Cardiovasc Imaging 2022;23(10):e479–e480.
19. Gao P, Yee R, Gula L, et al. Prediction of arrhythmic events in ischemic and dilated cardiomyopathy patients referred for implantable cardiac defibrillator: evaluation of multiple scar quantification measures for late gadolinium enhancement magnetic resonance imaging. Circ Cardiovasc Imaging 2012;5(4):448–456.
20. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
21. Iglovikov I, Shvets A. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. arXiv 1801.05746 [preprint]. https://arxiv.org/abs/1801.05746. Posted 2018. Accessed February 11, 2019.
22. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, 2015; 234–241.
23. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/. Published 2019. Accessed August 30, 2021.
24. Klem I, Heiberg E, Van Assche L, et al. Sources of variability in quantification of cardiovascular magnetic resonance infarct size - reproducibility among three core laboratories. J Cardiovasc Magn Reson 2017;19(1):62.
25. Moccia S, Banali R, Martini C, et al. Development and testing of a deep learning-based strategy for scar segmentation on CMR-LGE images. MAGMA 2019;32(2):187–195.
26. Tao Q, Piers SR, Lamb HJ, van der Geest RJ. Automated left ventricle segmentation in late gadolinium-enhanced MRI for objective myocardial scar assessment. J Magn Reson Imaging 2015;42(2):390–399.
27. Yue Q, Luo X, Ye Q, Xu L, Zhuang X. Cardiac Segmentation from LGE MRI Using Deep Neural Network Incorporating Shape and Spatial Priors. Springer Nature Switzerland AG 2019 D Shen et al (Eds): MICCAI 2019, LNCS 11765, pp 559–567. 2019. https://dl.acm.org/doi/abs/10.1007/978-3-030-32245-8_62.

28. Zabihollahy F, Rajchl M, White JA, Ukwatta E. Fully automated segmentation of left ventricular scar from 3D late gadolinium enhancement magnetic resonance imaging using a cascaded multi-planar U-Net (CMPU-Net). Med Phys 2020;47(4):1645–1655.
29. Roes SD, Borleffs CJ, van der Geest RJ, et al. Infarct tissue heterogeneity assessed with contrast-enhanced MRI predicts spontaneous ventricular arrhythmia in patients with ischemic cardiomyopathy and implantable cardioverter-defibrillator. Circ Cardiovasc Imaging 2009;2(3):183–190.
30. Alexandre J, Saloux E, Dugué AE, et al. Scar extent evaluated by late gadolinium enhancement CMR: a powerful predictor of long term appropriate ICD therapy in patients with coronary artery disease. J Cardiovasc Magn Reson 2013;15(1):12.
31. Rijnierse MT, van der Lingen ACJ, de Haan S, et al. Value of CMR and PET in Predicting Ventricular Arrhythmias in Ischemic Cardiomyopathy Patients Eligible for ICD. JACC Cardiovasc Imaging 2020;13(8):1755–1766.
32. Popescu DM, Shade JK, Lai C, et al. Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. Nat Cardiovasc Res 2022;1(4):334–343.
33. Wu KC. Sudden Cardiac Death Substrate Imaged by Magnetic Resonance Imaging: From Investigational Tool to Clinical Applications. Circ Cardiovasc Imaging 2017;10(7):e005461.
34. Demirel F, Adiyaman A, Timmer JR, et al. Myocardial scar characteristics based on cardiac magnetic resonance imaging is associated with ventricular tachyarrhythmia in patients with ischemic cardiomyopathy. Int J Cardiol 2014;177(2):392–399.
35. Guaricci AI, Masci PG, Muscogiuri G, et al. CarDiac magnEtic Resonance for prophylactic Implantable-cardioVerter defibrillAtor ThErapy in Non-Ischaemic dilated CardioMyopathy: an international Registry. Europace 2021;23(7):1072–1083.