DISSERTATION

Harmonization of confounding parameters to minimize induced biases on parametric T1 maps in cardiovascular magnetic resonance imaging

Harmonisierung von Einflussfaktoren zur Minimierung von induzierten Verzerrungseffekten bei parametrischen T1 Karten in der kardiovaskulären Magnetresonanztomographie

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Darian Steven Viezzer
aus Berlin

Erstbetreuung: Univ.-Prof. Dr. med. Jeanette Schulz-Menger

Datum der Promotion: 28.02.2025

Table of contents

List of tables

List of figures

List of equations

List of abbreviations

Abbreviation	Meaning
3D	three dimensional
AMY	patients with amyloidosis
BB	bounding box
BMI	body-mass-index
BPSP	best performing segmentation pipeline
CASEG	cascaded segmentation
CI	confidence interval
CMR	cardiovascular magnetic resonance imaging
CNN	convolutional neural network
COD	coefficient of determination
COV	coefficient of variation
CP	confounding parameter
CPIE	confounding parameter impact estimation
crinU	CASEG pipeline model with cropped image section and BB input
cropU	CASEG pipeline model with cropped image section input
cropU_A	CASEG pipeline model with cropped image section input (alternative)
CVD	cardiovascular disease
DB	database
DICOM	digital imaging and communications in medicine (data format)
DSC	Dice Similarity Coefficient
ECV	extra-cellular volume
ETR	extra-trees regression
FID	free induction decay
GUI	graphical user interface
HCM	patients with hypertrophic cardiomyopathy
HD	Hausdorff Distance
HR	heart rate
HTE	healthy test data
HTR	healthy training data
LAX	long axis
LCC	largest connected component
LGE	late gadolinium enhancement
LSVR	linear support vector regression
LV	left ventricle
LVH	patients with left ventricular hypertrophy
M	number of T1 maps
MAE	mean absolute error
MARISSA	Magnetic Resonance Imaging Software for Standardization
ME	mean error
mm	millimetre
MOLLI	modified Lock-Locker inversion recovery
MRI	magnetic resonance imaging
ms	milliseconds
N	number of subjects
ODA	object detection algorithm
r	Pearson correlation coefficient
refU	reference U-Net model
RFR	random-forest regression

RMSE	root mean squared error
ROC	receiver operating characteristic
ROI	region of interest
S	number of studies
SAX	short axis
SASHA	saturation-recovery single-shot acquisition
SCMR	society of cardiovascular magnetic resonance
ShMOLLI	shortend modified Lock-Locker inversion recovery
τ	Kendall correlation coefficient
T	Tesla
T1	longitudinal relaxation time
T1ρ	spin lattice relaxation time in the rotating frame
T2	transversal relaxation time
T2*	effective transversal relaxation time

Abstract

Background

Parametric T1 mapping is a quantitative method for myocardial tissue differentiation in cardiovascular magnetic resonance (CMR) that lacks comparability due to missing universal valid reference values. Confounding parameters (CP) impairing the myocardial T1 value quantification are described in literature. These CPs originates from subject, technologic and post-processing specific variations. The aim of this work was the evaluation of a post-hoc standardization approach for native parametric T1 maps within the self-developed Magnetic Resonance Imaging Software for Standardization (MARISSA). As quantitative measurements require a segmentation of the region of interest (ROI), a novel cascaded segmentation (CASEG) approach is additionally introduced as a necessary pre-processing step for MARISSA.

Methods

The proposed CASEG consisted of a bounding box (BB) prediction followed by a segmentation model. The BB was enlarged by a magnification factor of 1.5 to assure for full left ventricular coverage. Three CASEG pipelines were tested against a reference U-Net (refU): cropU, which used the enlarged BB image section, crinU, which used the enlarged image section and the original BB mask and cropU_A, which used the BB image section of a direct enlarged BB predictor. All models shared the same hyperparameters and were tested with respect to geometric and quantitative outcomes. The dataset included 403 subjects with 1080 native and 358 post-contrast T1 maps that were split into 75% training, 10% validation and 15% test data.

In MARISSA 214 healthy subjects (814 T1 maps) were used for training standardization models with respect to the CPs age, sex, scanner and sequence. Among the training dataset both sex, eleven scanners, eight sequences and an age distribution of 38±15years were available. Five adjustable standardization pipeline settings were optimized among 240 tested combinations by minimizing the coefficient of variation (COV) in a cohort of 40 healthy subjects (HTE, 156 T1 maps). The evaluated best performing standardization pipeline (BPSP) was then compared to 112 patients with a hypertrophic cardiomyopathy (HCM, 121 T1 maps) and 24 patients with an amyloidosis (AMY, 24 T1 maps).

Results

The Dice Similarity Coefficient as a measure of the geometric domain improved significantly for the test data in cropU, crinU and cropU_A (all around 80%) compared to refU (around 70%) while the mean absolute error improved only slightly without significance. The cropU represented the base segmentation in MARISSA. The BPSP halved the COV in the HTE to 6% while reaching a diagnostic sensitivity and specificity of 96%/92% between HTE and AMY, 72%/72% between HTE and HCM, and 88%/98% between HCM and AMY.

Conclusion

CASEG significantly improved the automatic segmentation in the geometric but not in the quantitative domain. MARISSA harmonized parametric T1 mapping values while maintaining the diagnostic accuracy for two dedicated patient groups.

Zusammenfassung

Hintergrund

Parametrische T1-Kartierung ist eine quantitative Methode zur myokardialen Gewebedifferenzierung in der kardiovaskulären Magnetresonanztomographie (CMR), der es an generischen Referenzwerten mangelt. In der Literatur werden Störparameter (CP) beschrieben, die den myokardialen T1-Wert beeinflussen. Diese CP entspringen Subjekt-, Technologie- und Nachverarbeitungs-spezifischen Variationen. Ziel dieser Arbeit war die Evaluierung eines nachgelagerten Standardisierungsansatzes für native T1-Karten innerhalb der selbstentwickelten Magnetic Resonance Software for Standardization (MARISSA). Da quantitative Messungen eine Segmentierung erfordern, wird zusätzlich eine neuartige kaskardierte Segmentierung (CASEG) als notwendige Vorverarbeitung für MARISSA eingeführt.

Methode

Die CASEG bestand aus einer Begrenzungsdetektion (BB) gefolgt von einem Segmentierungsmodell. Die BB wurde zur Abdeckung des gesamten linken Ventrikels um das 1.5-fache vergrößert. Drei CASEG-Modelle wurden gegen ein Referenz-U-Net (refU) getestet: cropU, das den vergrößerten BB-Bildausschnitt, crinU, das den vergrößerten Bildausschnitt und die BB-Maske, und cropU_A, das den BB-Bildausschnitt einer direkt vergrößerten BB verwendete. Alle Modelle wurden gleich eingestellt und hinsichtlich der ge-

ometrischen und quantitativen Ergebnisse getestet. Der Datensatz umfasste 403 Probanden mit 1080 nativen und 358 kontrastverstärkten T1-Karten, die in 75% Trainings-, 10% Validierungs- und 15% Testdaten aufgeteilt wurden.
In MARISSA wurden 214 gesunde Probanden (814 T1-Karten) für das Training von Standardisierungsmodellen in Bezug auf die CP Alter, Geschlecht, Scanner und Sequenz verwendet. Im Trainingsdatensatz waren beide Geschlechter, elf Scanner, acht Sequenzen und eine Altersverteilung von 38±15 Jahren vorhanden. Fünf anpassbare Einstellungen wurden unter 240 getesteten Kombinationen durch Minimierung des Variationskoeffizienten (COV) in einer Kohorte von 40 gesunden Probanden (HTE, 156 T1-Karten) optimiert. Das evaluierte beste Standardisierungsmodell (BPSP) wurde mit 112 Patienten mit hypertropher Kardiomyopathie (HCM, 121 T1-Karten) und 24 Patienten mit Amyloidose (AMY, 24 T1-Karten) verglichen.

Ergebnisse

Die Dice Metrik, als geometrisches Maß, verbesserte sich signifikant für die Testdaten in cropU, crinU und cropU_A (alle ca. 80%) im Vergleich zu refU (ca. 70%), während sich der mittlere absolute Fehler nur geringfügig verbesserte.
Die Basis-Segmentierung in MARISSA wurde durch cropU definiert. Das BPSP halbierte den COV im HTE auf 6 % und erreichte eine diagnostische Sensitivität und Spezifität von 96%/92% zwischen HTE und AMY, 72%/72% zwischen HTE und HCM und 88%/98% zwischen HCM und AMY.

Schlussfolgerung

CASEG verbesserte die automatische Segmentierung signifikant im geometrischen, aber nicht im quantitativen Bereich. MARISSA harmonisierte die Werte von parametrischen T1-Karten unter Beibehaltung der diagnostischen Genauigkeit für zwei dedizierte Patientengruppen.

1 Introduction

According to the most recent World Heart Report of 2023 by the World Heart Federation, cardiovascular diseases (CVDs) remain the leading cause for mortality affecting approximately one third of all global deaths[1]. Contemporary evidence in the research on assessing and treating CVDs open out into constantly updating guidelines[2,3]. Those guidelines recommend non-invasive imaging methods for the evaluation of individual heart characteristics[2,4]. The choice of the respective imaging modality depends on the questioning and accessibility[4,5]. Among those, cardiovascular magnetic resonance imaging (CMR) becomes increasingly important[6].

1.1 Cardiovascular Magnetic Resonance Imaging (CMR)

CMR is a specific use-case of magnetic resonance imaging (MRI), which reveals some unique opportunities compared to other imaging modalities as it provides high spatial resolution compared to echocardiography[7] while being untainted by ionizing radiation[8]. The physics of MRI is based on quantum mechanics of charged elements in a magnetic field[9–11]. The different acquisition techniques in MRI are commonly called sequences as they are based on a train of radiofrequency pulse transmission and resonating signal receipt[9–11]. By adapting the timing and other technical parameters the relaxation sensitivity and thus the contrast as well as the field of view are adjustable[9–11].

As CMR focuses on the heart, it suffers from two independent physiological motions during acquisition: breathing movement and cardiac motion[12]. The prior is compensated by either breath holding[13], respiratory gating[13] or a posteriori motion correction[14] while cardiac motion is handled by a cardiac gating signal that captures the same cardiac phase[15]. As the orientation of the heart in the human body is skewed[16], additional complexity for the definition of the imaging plane exist.

Nonetheless, CMR evolved to the state-of-the-art imaging modality for myocardial function, scarring and tissue characterization[17,18]. While volumetric quantification of the heart are accessible in any imaging modality[19], parametric mapping represents a unique quantification technique in CMR only[20,21].

1.2 Parametric T1 Mapping

The parametric mapping in CMR can either represent the T1, T2, T2* or T1ρ relaxation value of a tissue[22] and is calculated by a voxel-wise fitting of source images with different respective relaxation sensitivity[21]. Parametric T1, T2 and T2* mapping is already commercially available[23,24], increasingly used in clinical routine[25] and considered in current guidelines[2,5]. The integration of parametric T1 and T2 mapping into clinical CMR routine protocol revealed an improved diagnostic accuracy for CVD detection[26]. This work focuses solely on parametric T1 mapping as it showed among the different parametric mapping techniques the highest potential to differentiate healthy from various CVDs[20,22].
Parametric T1 maps are acquired either without (native) or after application of contrast agents (contrast enhanced). The combination of both, native and contrast enhanced, enables the calculation of the extra-cellular volume (ECV) map and complements the late gadolinium enhancement (LGE) quantification for myocardial fibrosis detection[27,28].
Across existing and ongoing developing T1 mapping sequences, the modified Lock-Locker inversion recovery (MOLLI)[29], its shortened version (ShMOLLI)[30] and the saturation recovery single-shot acquisition (SASHA)[31] emerged to the most common used T1 mapping sequences[32,33]. The MOLLI based T1 map shows high reproducibility in favour of accuracy while SASHA acts vice versa[33]. Furthermore, different schemes exist for MOLLI sequences which represent the number of acquired source images and in parenthesis the pause in-between, like in 5(3)3 b or 5(3)3 s, where b schemes clock by heartbeats while s schemes clock after seconds[33]. The ShMOLLI sequence is a MOLLI sequence with a 5(1)1(1)1 b scheme and an adapted fitting algorithm[30,34]. SASHA, in turn, is based on different saturation pulse times for the individual source image acquisitions[31].

1.3 Confounders

A CMR examination follows an imaging chain and with respect to parametric T1 mapping includes the acquisition with a dedicated sequence, the reconstruction of individual source images from the acquired signal including a motion correction that depends on the sequence variant and acquisition strategy[35,36], the generation of the parametric T1 map on the base of the motion corrected source images[37], a segmentation of the left ventricular myocardium as region of interest (ROI), whose output is used for the quantitative T1 statistics classification and finally its incorporation into the report[38] as abstractly shown in Figure 1.

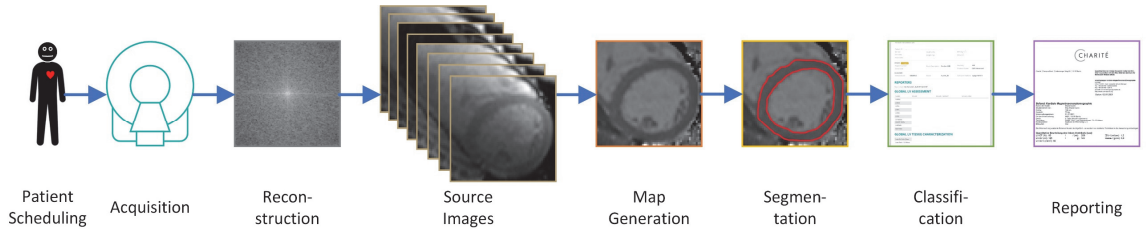

Figure 1: CMR imaging chain for parametric T1 mapping

Figure adapted from "Machine learning in cardiovascular magnetic resonance: basic concepts and applications" by Leiner et al.[38], 2019, under CC BY 4.0.

At any stage of the imaging chain potential confounders may induce a bias[22]. The confounding parameters (CPs) originate from subject specific sources, like age and sex[39,40], technical sources, like field strength[40], hardware[40] and sequence type[41,42] or assessment specific sources, like segmentation strategy[43] and experience of the operator[44,45]. The segmentation procedure represents a specialty among the CPs as it relies on manual contouring although even experienced operators show not neglectable uncertainty due to intra-observer variability[43]. Fully automatic segmentation procedures that are based on convolutional neural networks (CNNs) models and trained with expert's annotated data show in average a high geometrical concordance to expert segmentations while minimizing intra-observer variability[46–48]. However, individual outliers, high quantitative differences and a limited training dataset still require a manual inspection of those CNN model outcomes[46–48]. The impact of confounders undermines the accuracy and precision of quantified T1 values such that the reproducibility is limited to intra-institutional reference values only[20,49]. Consequently, standardization strategies are necessary to minimize the CPs induced biases.

1.4 Standardization

The aim of standardization in parametric T1 mapping is a harmonization of CPs in order to increase the reproducibility and thus the comparability but does not necessarily account for accuracy. Standardization strategies are either ad-hoc applied by suppressing the occurrence of a CP a priori or post-hoc deployed by subtracting out the impact of a CP a posteriori to its occurrence.

Among the ad-hoc standardization strategies, the usage of standard operating procedures following expert consensus as recommended by the Society for Cardiovascular Magnetic Resonance (SCMR) ensures a common understanding on the general application, handling and interpretation of parametric T1 maps[18,20].

A fixation of technical parameters, like field strength, MRI scanner vendor and sequence type achieved comparable T1 values with non-significant differences in a healthy volunteer cohort[50]. However, the strategy of constant CPs does not work out globally due to different accessibility, technical development and local preferences. For that reason, local reference values are recommended resulting in a limited comparability to other sites and the necessity of a local healthy reference cohort examination[20].

While ad-hoc strategies are preferred, their potential infeasibility makes post-hoc standardization methods necessary. The introduction of the z-score was the first attempt and is already mentioned in the clinical recommendations of the SCMR[20,51]. The score transforms parametric T1 mapping values into unitless values by offsetting against a local healthy reference cohort. Although the z-score already demonstrated its abilities[51], its drawback of a potential healthy reference cohort re-examination after a soft- or hardware change as well as its intrinsic standardization capabilities on solely technical CPs while subject specific CPs remain unaddressed may reasons for its restrained success.

At this point machine learning methods draw on by enabling the subtraction of individual CP biases as already shown for age and sex by linear regression models[39]. Those methods have a generalizability potential due to transfer capabilities onto other scanner-sequence combinations. However, the required constancy in the remaining CPs during the regression model training represents an obstacle and may have been yet the reason for the lack of such a comprehensive standardization approach for parametric T1 mapping.

1.5 Aims

In conclusion, native parametric T1 mapping shows a high potential to differentiate between healthy subjects and patients with CVDs while a stronger assertiveness in the clinical routine and its comparability still suffers from the lack of universal valid reference values[20,22]. This work aims to introduce and evaluate in a proof-of-concept a holistic regression model based post-hoc standardization pipeline for native parametric T1 mapping in CMR that is embedded in the extensible, self-developed prototype Magnetic Resonance Imaging Software for Standardization (MARISSA)[52,53]. MARISSA queues within the CMR imaging chain between segmentation and classification as a service providing software module and is not solely limited to CMR or T1 mapping. Since the segmentation plays an essential role and is an elusive CP, this work additionally includes a fully automated cascaded segmentation (CASEG) approach for parametric T1 maps[54] .

2 Methods

This retrospective work is approved by the local ethics committee of the Charité – Universitätsmedizin Berlin (EA1/253/21) and was supported by the BMBF (Bundesministerium für Bildung und Forschung) / DZHK (Deutsches Zentrum für Herz-Kreislauf-Forschung) via project FKZ81Z0100208. The used parametric T1 maps were inline generated at the scanner in the Digital Imaging and Communications in Medicine (DICOM)[55] format. Following the CMR imaging chain, the CASEG approach is described first, followed by an introduction of MARISSA. Both parts are structured with respect to Quality Assurance, Workflow, Datasets, Training, Testing and Implementation.

2.1 Cascaded Segmentation (CASEG)

The CASEG approach is an automatic segmentation method that consists of two consecutive steps. Firstly, the ROI is approximated with a rectangular bounding box (BB) by an object detection algorithm (ODA). Then, in a second step, the BB information is used to focus on the relevant image section within the subsequent segmentation model to detect the actual ROI. The following information on CASEG is retrieved from the publication by Viezzer et al.[54] and was exemplarily applied on CMR based parametric T1 maps.

2.1.1 Quality Assurance of CASEG

The aim of an automatic segmentation, like CASEG, is reaching common accord with manual expert contours. Consequently, geometric metrics represent quality features to measure the concordance between expected and achieved segmentation while quantitative metrics embody additional quality features as parametric T1 mapping is a quantitative method. The quality assurance metrics were applied in regard of the expert reference segmentation. The performance gain by incorporating the BB information was evaluated by testing against a standard automatic segmentation without the preliminary ODA.

2.1.1.1 Geometric Performance

The geometric performance of CASEGs were assessed by the Dice Similarity Coefficient[56] (DSC, Equation 1) and the Hausdorff Distance[56] (HD, Equation 2). The DSC, given in %, measures the consensus of two segmented areas. Whereas the HD, in millimeter (mm), represents a measure for the maximal local divergence.

$$DSC(A,B) = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

Equation 1: Calculation of the Dice Similarity Coefficient (DSC)

with A and B as the respective segmentation areas[56]

$$HD(A,B) = max\begin{Bmatrix} \underset{a \in cA}{max}\left\{\underset{b \in cB}{min}\{d(a,b)\}\right\} \\ \underset{b \in cB}{max}\left\{\underset{a \in cA}{min}\{d(a,b)\}\right\} \end{Bmatrix}$$

Equation 2: Calculation of the Hausdorff Distance

with A and B as the respective segmentation areas, cA and cB as the respective segmentation area contours and d represents the Euclidean distance[56]

2.1.1.2 Quantitative Performance

The quantitative performance of CASEGs were assessed by the mean error[57] (ME, Equation 3), mean absolute error[57] (MAE, Equation 4), root mean squared error[57] (RMSE, Equation 5) and the confidence interval[57] (CI, Equation 6) of mean T1 values. The ME, given in milliseconds (ms), represents the mean quantitative deviation and in combination with the MAE and RMSE, both given in ms, quantify the value dispersion. The CI, in turn, represents the value range that most probably capture the global mean T1 value[58].

$$ME(x,y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$$

Equation 3: Calculation of the mean error

with n as the number of values, y representing the expected and x the achieved values[57]

$$MAE(x,y) = \frac{1}{n}\sum_{i=1}^{n}|y_i - x_i|$$

Equation 4: Calculation of the mean absolute error

with n as the number of values, y representing the expected and x the achieved values[57]

$$RMSE(x,y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$$

Equation 5: Calculation of the root mean squared error

with n as the number of values, y representing the expected and x the achieved values[57]

$$CI(x) = \begin{bmatrix} \mu - \frac{c \cdot \sigma}{\sqrt{n}}\ (lower\ limit) \\ \mu + \frac{c \cdot \sigma}{\sqrt{n}}\ (upper\ limit) \end{bmatrix}$$

Equation 6: Calculation of the confidence interval

with x representing the values, μ representing the mean and σ representing the standard deviation of the values, n as the number of values and c as the confidence factor that depends on n and the significance level[57]

2.1.1.3 Statistical Analysis

The HD, DSC, ME and MAE were tested by a paired t-test and ANOVA in case of normal distribution according to a Shapiro-Wilk-test otherwise by a Friedman and Wilcoxon test with a significance level of $\alpha < 0.05$ in any case[59]. A higher DSC and lower HD, ME, MAE and RMSE were assumed with an improved performance. The equivalence margin derived from an intra-observer variability of native T1 maps is defined as the clinically acceptable deviation[43] and was used for testing the CIs. The CIs were Bonferoni corrected as the same test data was used for the CASEG pipeline and reference models[43].

2.1.2 Workflow in CASEG

The two steps in CASEG were implemented as two in sequence arranged CNNs. The primary acted as an ODA and predicted a binary BB for a rectangular ROI of the left ventricle (LV). Due to prediction uncertainties, the BB might miss tightly capturing the outer border of the LV and thus a BB enlarging magnification factor was used to assure for a whole LV coverage. The cropped image section by the enlarged BB was then fed into the secondary segmentation CNN to evaluate the actual LV ROI. Two different CASEG pipelines (cropU and crinU) were evaluated against a segmentation without a preliminary ODA (refU) as shown in Figure 2.

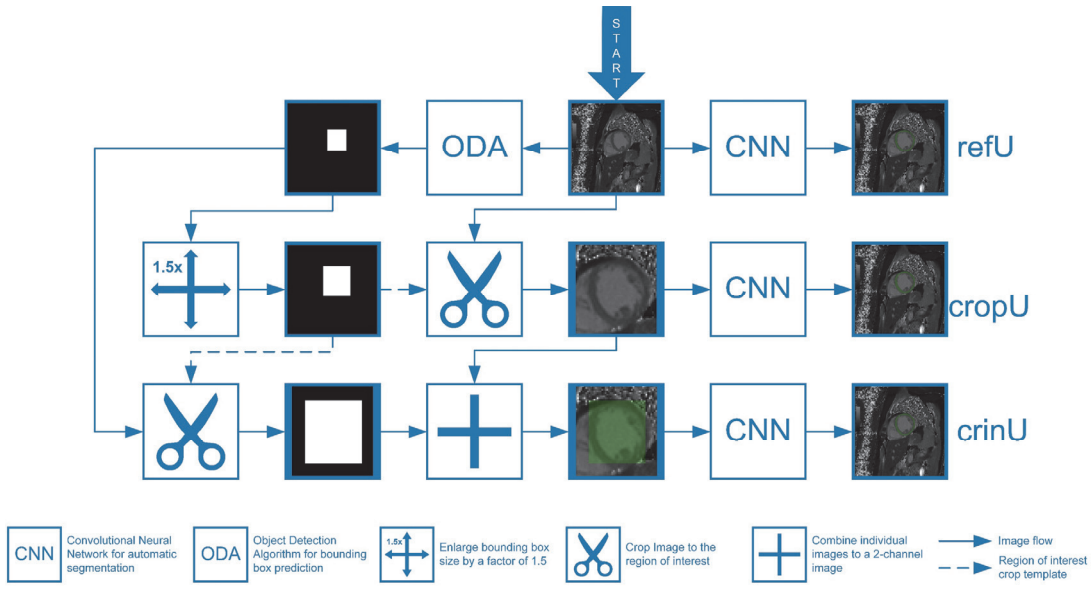

Figure 2: CASEG workflow

The reference segmentation by a convolutional neural network (CNN) without a preliminary object detection algorithm (ODA) for the heart (refU) directly predicts the segmentation on the original image. In contrast to this, cropU focuses solely on the image section that is defined by the enlarged predicted bounding box while crinU works equivalent but has an additional second input channel reflecting the original predicted bounding box mask. Figure retrieved from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0.

Both CASEG pipelines followed the same scheme, but crinU had a second input channel in the segmentation CNN to include the original predicted BB mask. An alternative cropU version (cropU_A) was evaluated with an ODA CNN that directly predicted an enlarged BB instead of a tight one that needed an enlarging processing.

2.1.2.1 Bounding Box (BB) Detection

The CNN based BB detection outputs an array of the same size as the input image with pixel values of floating point-numbers between zero and one. A threshold of 0.5 turned the output into a binary image with relevant (one) and non-relevant (zero) parts. The largest connected component (LCC) was identified and the BB edges were evaluated by the minimum and maximum indices of the LCC along both image axes. This BB was then enlarged by a magnification factor that was evaluated as the minimal factor for full coverage of the LV in the dedicated test dataset.

2.1.2.2 Automatic Segmentation

The automatic segmentation in the secondary part of the CASEG was also implemented as a CNN. All CNNs were in turn implemented as U-Nets[60] as shown in Figure 3.

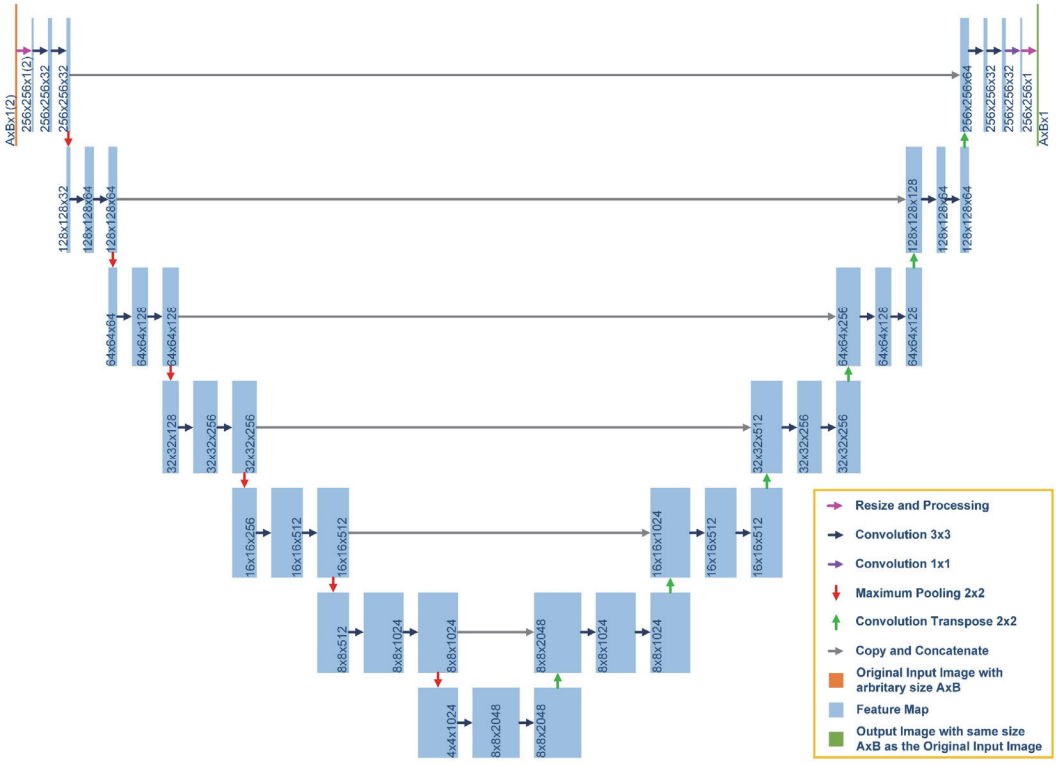

Figure 3: Convolutional Neural Network (CNN) structure of the U-Nets as implemented in CASEG

Figure adapted from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0.

For an unadulterated evaluation of the BB performance gain, all U-Nets shared the same set of hyperparameters with the only exception of the segmentation CNN in crinU which required two input channels. The U-Nets had 27 layers and accordingly six skip connections and an input size of 256x256x1 respectively 256x256x2 and an output size of 256x256x1. The input images were resized and interpolated accordingly while the output was scaled back to fit the original image size. A conversion of binary pixel masks into vectorized contour objects, application of geometric transformation in the vectorized domain and then back-transformation into binary pixel masks by rasterization assured for a lossless propagation.

The chosen hyperparameters Adam optimizer[61] with a clipnorm of 0.001, log-cosh-dice loss function[62], 0.001 learning rate and reduce-on-plateau learning rate adaption[63] were pre-evaluated as the best combination among the refU training for the tested loss functions: binary-cross-entropy and log-cosh-dice[62], learning rates: 0.01 and 0.001 and learning rate adaptions: constant, exponential, linear, reduce-on-plateau and triangle[63,64]. The training setup included a batch size of ten and 1000 epochs with an early stopping scheme[63,65] after 50 epochs of no improvement with respect to the DSC in a separate validation dataset as well as a halving of the initial learning rate of 0.001 after 25 epochs of no improvement. A threshold of 0.5 was applied on the output in order to retrieve a binary segmentation mask.

2.1.3 Datasets for CASEG

Data from eleven published[50,66–68] and on-going[69] studies was used with in total N=403 participants (97 healthy volunteers and 306 patients) and M=1438 parametric T1 maps (1080 native and 358 contrast enhanced). All T1 maps had manual reference segmentations by experienced readers using the software cvi42 (Circle Cardiovascular Imaging, Calgary, Canada) and were either generated on a 1.5T AvantoFit, a 3T SkyraFit or a 3T PrismaFit clinical magnetic resonance imaging scanner (all Siemens Healthcare, Erlangen, Germany). The native T1 maps were based on a 5(3)3 b MOLLI sequence while the contrast enhanced T1 maps were based on a 4(1)3(1)2 b MOLLI sequence. The split into 75 % training, 10 % validation and 15 % test data were randomly performed for each study set of the CASEG dataset as shown in Table 1. During training and validation, any slice location was used while testing was performed on midventricular and basal slices only as recommended by the SCMR[20].

Table 1: CASEG dataset overview

	Training	Validation	Testing	Total
subjects (N)	313	35	55 (55)	403
native T1 maps	849	91	140 (106)	1080
contrast enhanced T1 maps	286	27	45 (33)	358
total T1 maps (M)	1135	118	185 (139)	1438

As the testing included midventricular and basal slices only, the numbers in (parenthesis) represent the actual amount of used test data. Table retrieved from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0.

2.1.4 Training of CASEG

During training of the CNNs the training dataset was used while the validation dataset acted as an unknown dataset to validate the performance with respect to the DSC. This reduced the risk of overfitting and thus improved the generalization of the CNN. The training dataset was randomly augmented[70] with adjustments in brightness and contrast, noise by blurring, Gaussian random as well as salt and pepper noise and transformations by rotation, mirroring, and axis down-sampling. A cropping of the input image was randomly performed for the refU CNN and ODA CNN, while in cropU, crinU and cropU_A, the detected BB was randomly shifted and resized by up to five pixels. In respectively five percent of the cases the optimal BB and no detected BB were assumed to include potential ODA uncertainties. In case of a missing BB cropU cannot focus on the ROI and consequently acts like refU, while crinU uses a zero-valued secondary input array. The BB enlargement was limited to the image boundaries. Each input image channel was normalized by the brightest image value to floating-point values between zero and one.

2.1.5 Testing of CASEG

The information content of an image, calculated by the ratio of relevant pixels to the total number of pixels, was statistically tested by a t-test in case of normal distribution according to the Shapiro-Wilk-test otherwise by a Wilcoxon test with a significance level of $\alpha < 0.05$[59] between the image section of the enlarged BB and the original image. The performance of the individual CASEG pipelines were evaluated against refU and against each other according to the prior defined quality assurance of CASEG. A correlation and a Bland-Altman plot were provided to visualize concordance and the limits of agreement[71]. The correlation plot includes a linear regression, the Pearson Correlation (testing for linearity) and Kendall's Tau (testing for rank-order stability or monotony)[59]. While the correlation coefficient values were assumed as weak if smaller than 0.35, moderate if up to 0.67, strong if up to 0.90 and very strong if above, the coefficient of determination

(COD, squared Pearson correlation coefficient) was used to support the interpretation of the findings by representing the shared variance across two measurements[72]. As native and contrast enhanced T1 times have different scales, the testing results of the CASEG pipelines were separated for native and contrast enhanced datasets.

2.1.6 Implementation of CASEG

The U-Nets were implemented in Tensorflow[73], statistics were calculated with scipy[74], plots were generated with matplotlib[75] and everything was cast into Python (Version 3.8, Python Software Foundation, Beaverton, USA). Table 2 shows an overview of all necessary Python site-packages for CASEG. The source code was made publicly available via GitHub: *https://github.com/DSV-CUB/CASEG*

Table 2: List of Python site-packages for the CASEG implementation

Package	**Version**
GDAL[76]	3.4.1
geopy[77]	2.2.0
h5py[78]	2.10.0
keras[63]	2.7.0
matplotlib[75]	3.5.2
mock[79]	4.0.2
numpy[80]	1.21.5
opencv-python[81]	4.4.0.42
openpyxl[82]	3.0.9
pandas[83]	1.3.4
Pillow[84]	7.2.0
pip[85]	20.1.1
polyline[86]	1.4.0
pydicom[87]	2.2.2
rasterio[88]	1.2.10
scikit-fuzzy[89]	0.4.2
scikit-image[90]	0.18.1
scikit-learn[91]	1.0.2
scipy[74]	1.4.1
shapely[92]	1.8.0
statsmodels[93]	0.13.0
tensorflow[73]	2.7.0
tensorflow-gpu[73]	2.7.0
XlsxWriter[94]	3.0.1
xlwings[95]	0.25.0

Table retrieved from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0.

2.2 Magnetic Resonance Imaging Software for Standardization (MARISSA)

The MARISSA provides regression model based post-hoc standardization of quantitative measurements in medical imaging as a service. A standardization is setup as a pipeline and transforms quantitative values of arbitrary CP settings into values of a pre-defined and fixed reference CP environment. The prototyped software includes a graphical user interface (GUI) for usability and a modularized structure for extensibility. The following information on the MARISSA are based on the publication by Viezzer et al.[52] and were evaluated as a proof-of-concept on CMR based native parametric T1 maps.

2.2.1 Quality Assurance of MARISSA

The aim of a standardization approach for quantitative measurements, as enabled with the MARISSA, is reaching comparability in the quantified values. Therefore, the quality assurance focuses on the quantitative domain of the segmented ROI with evaluations on the intra-cohort scale to minimize variation within the same cohort and on the inter-cohort scale to maximize the differentiability between cohorts.

2.2.1.1 Quantitative Intra-Cohort Performance

The quantitative intra-cohort performance was assessed by the coefficient of variation (COV, Equation 7) as it represents a normalized and thus unitless, measure of the dispersion within a group[58]. A minimization of the COV was assumed with a successful standardization performance.

$$COV = \frac{\sigma}{\mu}$$

Equation 7: Calculation of the coefficient of variation with σ as standard deviation and μ as mean[58]

2.2.1.2 Quantitative Inter-Cohort Performance

The quantitative inter-cohort performance was evaluated by the receiver operator characteristics (ROC) analysis for differentiation between healthy subjects and patients with various CVDs[96]. The ROC includes the evaluation of an optimal limit with respective sensitivity and specificity[96]. The diagnostic accuracy after standardization was assumed as evident if the sum of both, sensitivity and specificity, exceeded 150%[97] and were in the range of or above an intra-scanner-intra-sequence ROC analysis.

2.2.1.3 Statistical Analysis

The T1 values of the considered cohorts in MARISSA were statistically tested by an independent t-test in combination with an ANOVA test in case of normal distribution according to the Shapiro-Wilk-test otherwise with the Mann-Whitney-U and Kruskal-Wallis test, all with a significance level of $\alpha < 0.05$[59]. CIs were plotted against the CI of datasets that captured the reference CP environment to check if healthy subjects stay within while patients with CVDs clearly stand out.

2.2.2 Workflow in MARISSA

The core of the MARISSA is the SQLite[98] database MARISSA DB that is automatically created for each user defined project within the software. A detailed manual on the usage of MARISSA was provided with the publication[52]. The MARISSA DB contains relational connected tables separated in an active and a passive site as shown in Figure 4. The active site consists of all data and information the user actively manipulates by defining, editing and setting. This includes the import of DICOM and segmentation data as well as the definition of standardization pipelines and CPs from available DICOM tags. The passive site is automatically filled during training and contains information about the trained regression models including a track of the training datasets. For standardization, the respective trained models are read from the MARISSA DB and applied onto the dataset.

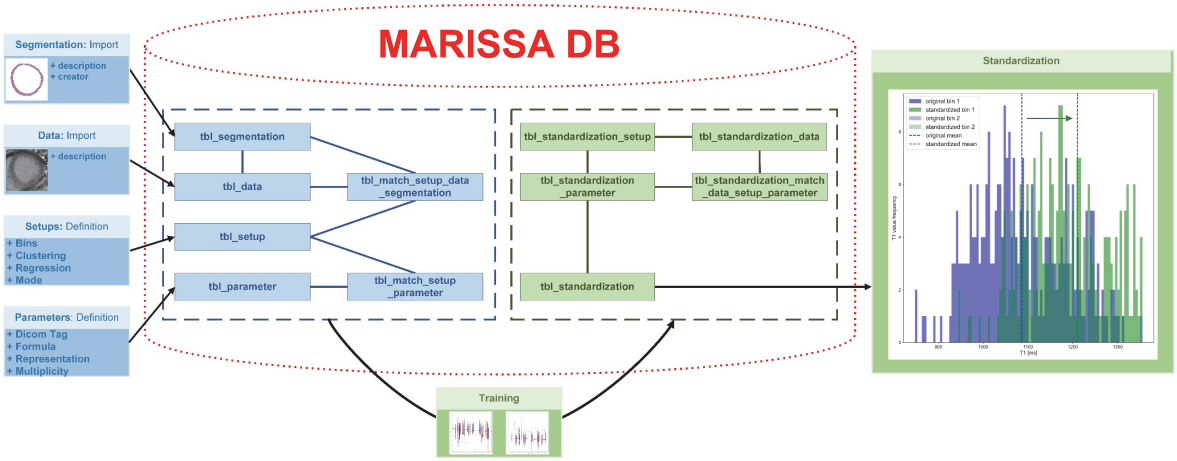

Figure 4: MARISSA database (DB) structure

The blue tables tbl_segmentation, tbl_data, tbl_setup, tbl_parameter, tbl_match_setup_data_segmentation and tbl_match_setup_parameter denote the active site where the user interacts actively by importing, defining, editing and manipulating data whereas the training information is copied into the separate green tables tbl_standardization_setup, tbl_standardization_data, tbl_standardization_parameter, tbl_standardization_match_data_setup_parameter and tbl_standardization. The green tables are automatically filled during training of a standardization pipeline and contain all necessary information and data to trace back fitted regression models. Figure retrieved from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

2.2.2.1 Regression Model

The regression models were trained to estimate the induced biases of CPs and therefore labelled as CP impact estimators (CPIEs). Due to the lack of either accuracy or precision in T1 mapping sequences[33], the true T1 value is unknown and thus the induced biases are only relatively calculatable. For that reason, the reference CP environment defines for each CP a zero-bias assumed reference value. The induced bias is the difference between a measured T1 value and the average T1 value across all datasets that capture the reference CP values (Equation 8 and Equation 9). The regression model only uses CP values as input and outputs the estimated bias in either absolute or relative [%] unit of measures.

$$bias_{abs} = y_{measured} - y_{reference} \; with \; y_{reference} = \frac{1}{|RD| \cdot |ROI|} \sum_{i=1}^{|RD|} \sum_{j=1}^{|ROI|} y_j$$

Equation 8: Calculation of the absolute bias

Subtraction of the reference quantitative value from the measured quantitative value, where the reference quantitative value equals the mean quantified values in the segmented region of interest (ROI) across all relevant datasets (RD) whose confounding parameters (CPs) capture the reference CP environment

$$bias_{\%} = 100\% \cdot \frac{bias_{abs}}{y_{reference}} \; with \; y_{reference} = \frac{1}{|RD| \cdot |ROI|} \sum_{i=1}^{|RD|} \sum_{j=1}^{|ROI|} y_j$$

Equation 9: Calculation of the relative bias

Division of the reference quantitative value from the absolute bias where the reference quantitative value equals the mean quantified values in the segmented region of interest (ROI) across all relevant datasets (RD) whose confounding parameters (CPs) capture the reference CP environment

2.2.2.2 Standardization

The standardization subtracts the CPIE results from the quantified values and returns the standardized values (Equation 10 and Equation 11). As a regression model only approximates the bias, the output is labelled as estimated CP error (ecpe) to discriminate from the true underlying bias. While numerical CPs are continuous and therefore also standardizable for unknown CP values during training, categorical CPs can be standardized for CP values only that were included in the training of the respective regression models.

$$y_{standardized} = y_{measured} - \sum_{i=1}^{|CP|} ecpe_{abs,i} \; with \; ecpe_{abs} \approx bias_{abs}$$

Equation 10: Standardization calculation for absolute confounding parameter (CP) impact estimations with ecpe representing the estimated confounding parameter error in absolute values

$$y_{standardized} = \frac{y_{measured}}{\left(100\% + \sum_{i=1}^{|CP|} ecpe_{\%,i}\right)} \; with \; ecpe_{\%} \approx bias_{\%}$$

Equation 11: Standardization calculation for relative confounding parameter (CP) impact estimations with ecpe representing the estimated confounding parameter error in relative values (%)

2.2.3 Datasets for MARISSA

Midventricular, native T1 mapping data from published[50,51,99–102] and on-going studies was used with in total N=254 healthy volunteers (Healthy), N=112 patients with left ventricular hypertrophy (LVH) including hypertrophic cardiomyopathy (HCM) and N=24 patients with amyloidosis (AMY) resulting in M=970, M=121 and M=24 midventricular native T1 maps in the respective cohorts as shown in
Figure 5. The surplus of T1 maps compared to the number of subjects originates from multiple measurements of some subjects under different scanner or sequence conditions. The two patient cohorts were chosen because AMY are expected for having significantly higher, barely overlapping T1 value distributions compared to Healthy while HCM have significant higher but largely overlapping T1 value distributions[103–105].
As a proof-of-concept for the post-hoc standardization of parametric T1 maps, the CPs age[39,40], sex[39,40], scanner model[40] and sequence variant[41,42] were chosen from literature to standardize for. The respective overview on the dataset composition on each CP is given in
Figure 5 as well. The CP reference environment was set to 18 years, male, 3.0T Siemens Verio [syngo MR B17] and MOLLI 5(3)3 b respectively for age, sex, scanner and sequence.
All T1 maps received an automated cropU CASEG segmentation[54] as well as a research deep learning segmentation that was provided by Siemens Healthcare (version 21 hotfix, Siemens Healthcare GmbH, Erlangen, Germany). Both segmentations were supervised by an expert who chose the segmentation for the individual map among either segmentation model or the intersection of both. A manual segmentation was performed in cvi42 (Circle Cardiovascular Imaging, Calgary, Canada) in case of failed segmentations among both segmentation models.
While patients suffering from the same CVD share similar myocardial degenerations, the state and phase of that CVD may influence the myocardial T1 value[106,107]. In order to omit CVD specific influences, only the Healthy cohort was used during the regression model trainings with a split into 85% training (HTR) and 15% testing (HTE) data on the base of the number of subjects per study. Consequently, N=214 subjects (with 100 males and 114 females, an age of 38.46±15.20 years and M=814) were in the HTR and N=40 (with 18 males and 22 females, an age of 39.50±15.89 years and M=156) were in the HTE group.

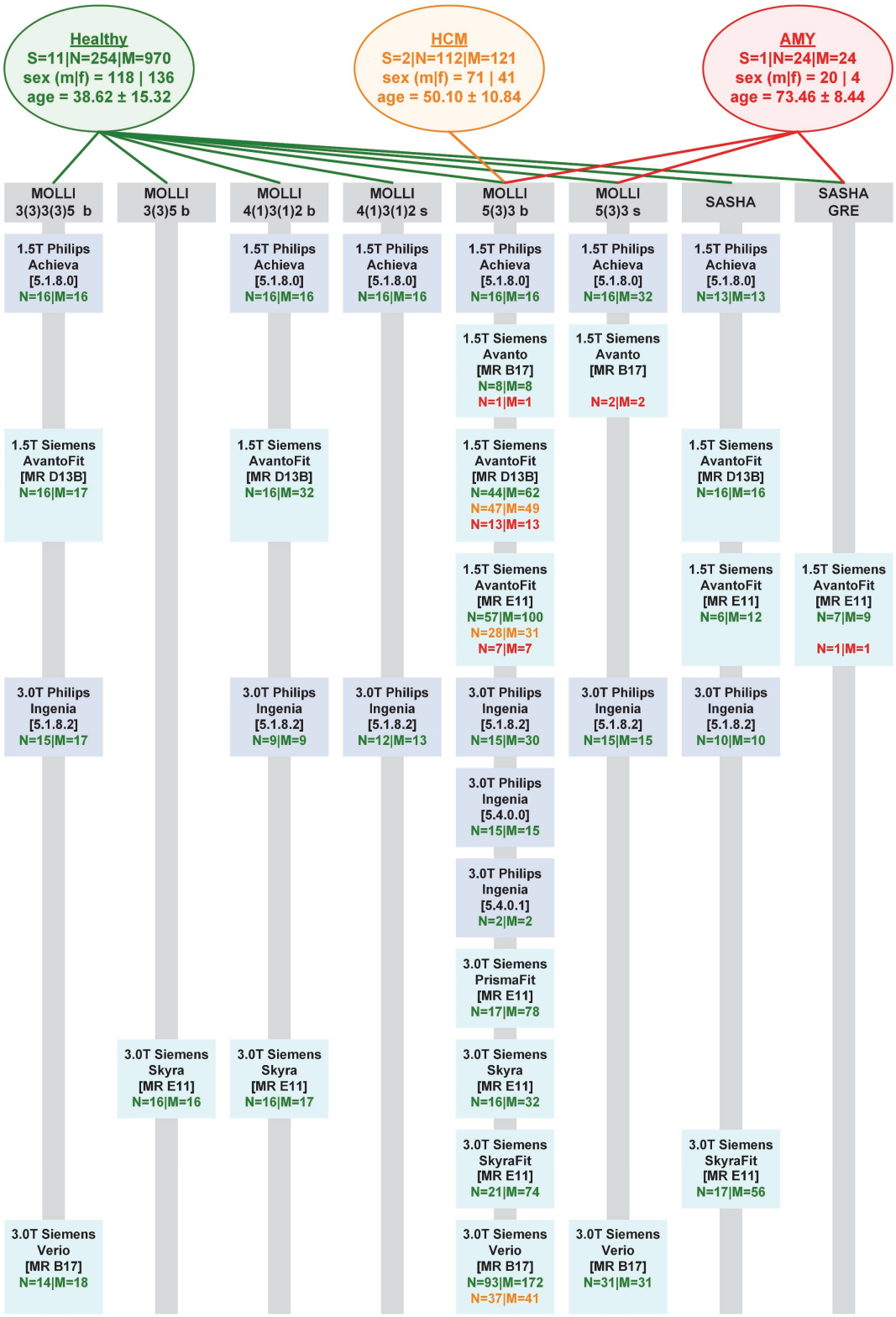

Figure 5: MARISSA dataset overview

A detailed breakdown on the respective scanner-sequence combinations with S: number of studies, N: number of subjects, M: number of T1 maps, m: male and f: female; age is given in mean ± standard deviation; green denotes healthy, orange denotes patients with hypertrophic cardiomyopathy and red denotes patients with amyloidosis. Figure retrieved from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

2.2.4 Training in MARISSA

A standardization pipeline in MARISSA is defined by the setting of the regression-type, the unit of measure, the mode and, if applicable, the clustering-type including the number of bins with current options shown in Table 3.

Table 3: Standardization pipeline settings

Setting	Options
regression-type	extra-trees (ETR) linear linear support vector (LSVR) random-forest (RFR)
unit of measure	absolute relative
mode	cascaded ensemble individual
number of clusters (bins)	1 to minimum number of segmented pixels
cluster-type	agglomerative average agglomerative complete agglomerative single agglomerative ward equal distant equal size gaussian mixture k-means

Table adapted from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

While the linear regression is the most basic and commonly used regression type[108], the linear support vector regression[109] (LSVR) is an alternative linear model. Additionally, random forest[110] regression (RFR) and extra-trees[111] regression (ETR) were evaluated in order to reflect potential non-linear CP bias relationships. The estimated bias was either calculated in absolute [ms] or relative [%] values. In absolute mode the standardization shifts the T1 values while maintaining the intrinsic spread whereas the relative mode weights the estimated bias according to the measured T1 value. The mode represents the standardization pipeline strategy. The individual and cascaded mode include distinct regression models for each CP. While the individual mode trains each regression model individually by constraining constancy in all CPs except the one of interest, the cascaded mode iteratively standardizes the training dataset with respect to each trained CPIE. The ensemble mode, in turn, trains a single regression model that includes all CPs at once and is expected to account for cross-dependencies in the CPs that are not considerable in the individual and cascaded mode. This mode converts categorical CPs into integer

numbered classes. A bin larger one enabled the clustering of T1 values in order to account for potential, currently unknown, cross-dependencies between acquired T1 values and CP values with individual CPIEs for each cluster and CP. The agglomeratives, gaussian mixture and k-means clustering algorithms are common used ones[91]. The equal distant and equal size clustering were self-implemented and both rely on sorting the T1 values from low to high. The equal distant clustering considers the minimum and maximum occurring T1 value and creates clusters of equal value ranges, whereas the equal size clustering allocates the same amount of T1 values into each cluster.

In this work, the best performing standardization pipeline (BPSP) was evaluated in two steps. First, the 24 standardization pipeline combinations of regression-type, unit of measure and mode without clustering were trained on HTR and evaluated for the COV on HTE. The three best performing pipeline settings with the lowest COV were then trained for two to ten bins among the eight different cluster-types. Although the exceeding of ten bins is conceivable, it is not recommended due to potential cluster sized being too small. Including these additional 216 standardization pipelines, the BPSP was identified across all 240 trained pipelines with the lowest COV in the HTE.

2.2.5 Testing of MARISSA

The evaluation of the BPSP during training already covers the intra-cohort quality assurance among the HTE. Therefore, the testing includes the intra-cohort analysis for the patient groups and the subsequent inter-cohort quality assurance by evaluating the diagnostic performance of the BPSP. The ROCs of the standardized cohorts were compared to intra-scanner-intra-sequence diagnostics. Additional boxplots and CIs were used to visualize overlapping value ranges between the cohorts accompanied by statistical testing. A progression plot was used to illustrate the individual CP impact. Finally, an additional sub-analysis of eight individuals of the HTE, who received multiple acquisitions under different scanner-sequence-combinations, was performed in regard of the COV and the absolute value spread among the mean T1 values of the individual measurements.

2.2.6 Implementation of MARISSA

The MARISSA was fully implemented in Python (Version 3.8, Python Software Foundation, Beaverton, USA) with a SQLite database backend and a PyQt5[112,113] GUI. The necessary Python site-packages are listed in Table 4. The software was made publicly available via GitHub: *https://github.com/DSV-CUB/MARISSA*[53].

Table 4: List of Python site-packages for the MARISSA implementation

Package	Version
fnv[114]	0.2.0
GDAL[76]	3.4.3
matplotlib[75]	3.5.2
mock[79]	4.0.2
numpy[80]	1.23.5
opencv-python[81]	4.6.0.66
pandas[83]	1.5.2
Pillow[84]	7.2.0
pydicom[87]	2.2.2
PyQt5[112]	5.15.4
PyQt5Designer[113]	5.14.1
rasterio[88]	1.2.10
rpy2[115]	3.5.4
scipy[74]	1.10.1
seaborn[116]	0.12.2
shapely[92]	2.0.0
scikit-image[90]	0.18.1
scikit-learn[91]	0.23.2
sqlalchemy[117]	1.4.41
tensorflow[73]	2.7.0
XlsxWriter[94]	3.0.1

Table adapted from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

The software architecture scheme of MARISSA is depicted in Figure 6 and consists of the three main components: gui, modules and toolbox. The designs sub-component in gui contains all PyQt5 design files and the Python files in the gui component connect the controls, like buttons, with executable Python procedures. The images directory in the gui component contains portable network graphic images that are displayed within the GUI. The toolbox component contains a collection of basic functions in the tools sub-component and basic classes in the creators sub-component. Finally, the modules component is the core of MARISSA and contains all the logics. The clustering sub-component has a dedicated Python file with the same class structure for each clustering algorithm. Following this structure, MARISSA is easily extensible by novel clustering algorithms. The GUI automatically detects available clustering algorithms from this component directory. Equivalent to the clustering sub-component, the regression sub-component contains individual Python files for each regression-type with a standardized structure such that MARISSA is also easily extensible for other regression methods. All current regression

models were implemented in the default setting with scikit-learn[91]. RFR and ETR were solely adapted to an increased number of estimator trees of 1000 that enabled an improved performance without an overfitting risk due to its intrinsic structure[110].

As standardization is applicable on accessible CPs only, CPs are defined by available DICOM tags. Standard DICOM tags are included in MARISSA by default while custom CPs are enabled by processing DICOM tags. The value representation according to the DICOM standard defines a CP as either numerical or categorical parameter.

After application of a standardization pipeline on a dataset, the original DICOM data, an Excel table that includes the CP values and the segmented T1 values with their respective progression during standardization as well as a MARISSADATA file and a progression plot are exported. The MARISSADATA file makes the information from the Excel file directly available for further processing in Python by a pickled dictionary. Further detailed information about the usage of MARISSA was provided in the user manual[52].

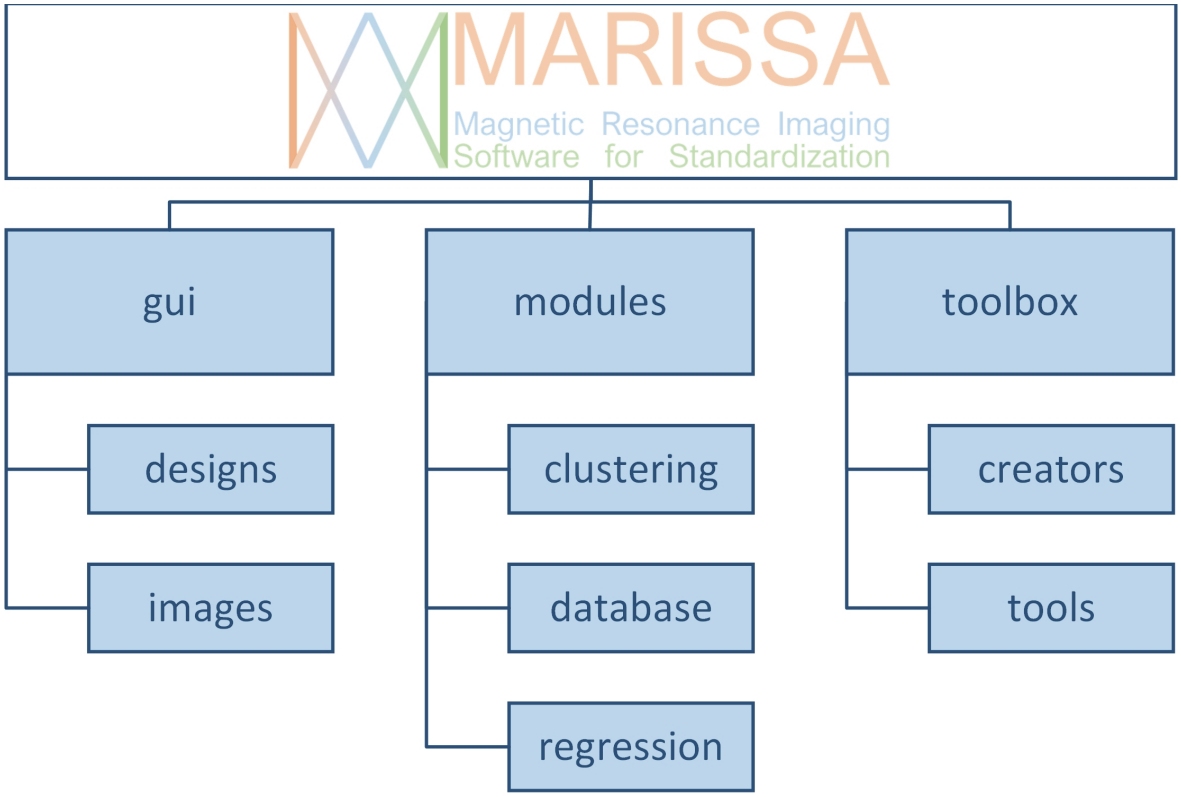

Figure 6: MARISSA architecture

MARISSA consists of the three main components: gui, modules and toolbox. The gui module contains everything for the graphical user interface including the PyQt5 design files and Python files for the control procedures. The toolbox contains general functions and re-usable classes. The modules component contains all the logic and calculations that define the core of MARISSA. Figure adapted from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0..

3 Results

The automatic CASEG as well as the MARISSA and the included standardization pipelines were successfully implemented and evaluated. After demonstrating the results of the CASEG, its usage in MARISSA is shown and finally the MARISSA results are provided. Numeric values are usually shown as mean ± standard deviation with their specific unit of measure.

3.1 Cascaded Segmentation (CASEG)

The CASEG was evaluated for the three segmentation pipelines: cropU, crinU and cropU_A against the reference segmentation model refU. The following results are retrieved from the publication by Viezzer et al.[54].

3.1.1 Training of CASEG

The training for the ODA and segmentation models were performed separately. The ODA identified after training a BB for the LV in all CASEG test datasets with a DSC of 93.09±2.13% and 91.99±2.80% and a HD of 3.95±1.07mm and 4.42±2.40mm for native and contrast enhanced T1 maps respectively. Figure 7 exemplarily shows the best and worst predicted BBs with respect to the DSC and HD metrics among the test data.
A testing of cropU and crinU among BB magnification factors from 1.0 to 2.5 in 0.1 steps revealed a magnification factor of at least 1.3 whereas a factor of 1.44 was necessary by calculation. Hence, a rounded-up magnification factor of 1.5 was used in cropU and crinU, while the ODA in cropU_A was trained to directly predict the 1.5 times enlarged BB. The enlarged BB ROI in turn revealed a significant higher ratio of relevant pixels over total pixels of 10.38±3.27% and 10.76±4.22% compared to the original image with 0.71±0.35% and 0.62±0.21% for native and contrast enhanced test datasets respectively. Optimal BBs would increase this ratio to a maximum of 20.12±6.67% and 19.78±6.00 % respectively.
The early stopping scheme was triggered in all trained models. The number of trained epochs were 191, 219, 168 and 155 resulting in a training data DSC of 81.04%, 83.62%, 81.58% and 81.57% as well as a validation DSC of 83.01%, 82.82%, 82.31% and 82.77% for refU, cropU, crinU and cropU_A respectively.

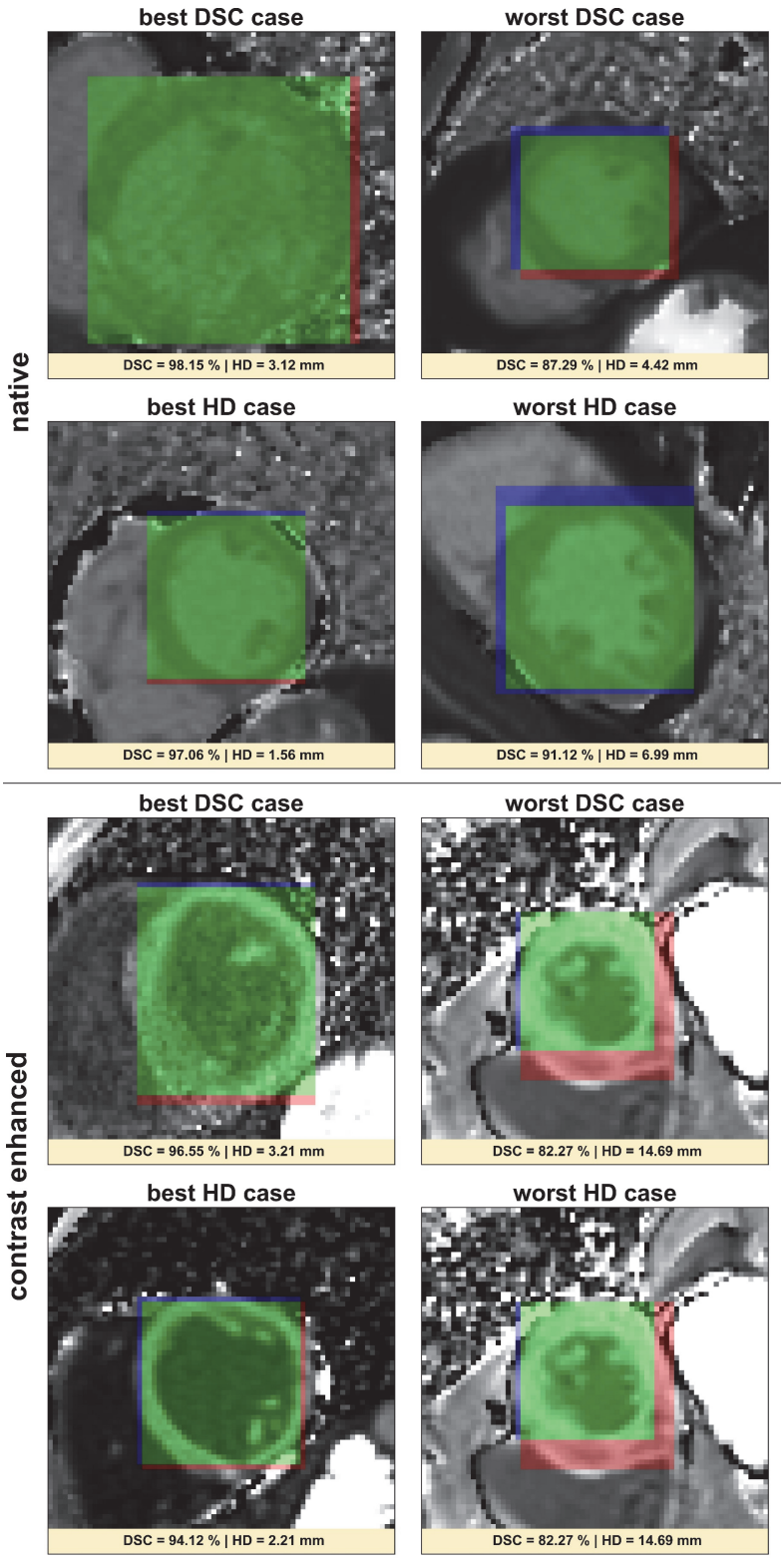

Figure 7: CASEG results of the object detection algorithm (ODA) as exemplary plots

The best and worst cases with respect to the Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics are displayed. Green denotes consensus with expert segmentation, red denotes false positive and blue denotes false negative segmentation. Figure adapted from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0.

3.1.2 Testing of CASEG

The complete geometric and quantitative results are provided in Table 5, separated for native and contrast enhanced T1 maps. Segmentation results for refU, cropU, crinU and cropU_A are exemplarily shown for a good, bad and with respect to refU improving case in Figure 8.

Table 5: CASEG numeric results for refU, cropU, crinU and cropU_A in the geometric and quantitative domain

	Metric		refU	cropU	crinU	cropU_A
native	Geo-metric	*DSC (%)*	72.79±8.08	81.06±5.57*	81.22±5.52*	81.13±5.83*
		HD (mm)	3.74±1.37	2.95±1.06*	3.01±1.20*	2.98±1.16*
	Quantitative	*ME (ms)*	-7.22±17.19	-6.00±14.67	-5.24±16.40*	-3.88±16.10*
		MAE (ms)	14.22±12.06	11.94±10.43	12.45±11.89	12.45±11.89
		RMSE (ms)	18.64	15.85	17.22	16.56
		CI (ms)	-11.38 / -3.05	-9.56 / -2.44	-9.22 / -1.26	-7.79 / 0.02
		COV (%)	2.38	2.45	3.13	4.15
		r (Pearson)	0.97*	0.97*	0.97*	0.97*
		COD (%)	94.09	94.09	94.09	94.09
		τ (Kendall)	0.80*	0.83*	0.82*	0.83*
contrast enhanced	Geo-metric	*DSC (%)*	71.41±8.54	80.70±10.31*	79.18±10.20*	80.15±10.21*
		HD (mm)	3.83±1.44	3.08±1.72*	3.35±1.90*	3.27±2.05*
	Quantitative	*ME (ms)*	5.23±8.14	4.45±8.39	5.17±7.27	4.57±7.85
		MAE (ms)	5.89±7.67	5.32±7.87	6.07±6.54	5.07±7.53
		RMSE (ms)	9.67	9.50	8.92	9.08
		CI (ms)	1.70 / 8.77	0.81 / 8.10	2.01 / 8.34	1.15 / 7.98
		COV (%)	1.56	1.89	1.41	1.72
		r (Pearson)	0.98*	0.98*	0.98*	0.98*
		COD (%)	96.04	96.04	96.04	96.04
		τ (Kendall)	0.91*	0.91*	0.92*	0.94*

*The values are separated for native and contrast enhanced T1 maps. * denotes statistical significance with a significance level of α < 0.05; DSC: Dice Similarity Coefficient; HD: Hausdorff Distance; ME: Mean Error; MAE: Mean Absolute Error, RMSE: Root-Mean-Squared Error; CI: confidence interval; COV: coefficient of variation; r: Pearson's correlation coefficient; COD: Coefficient of Determination; τ: Kendall's Tau coefficient; values are given as mean±standard deviation. Table retrieved from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0*

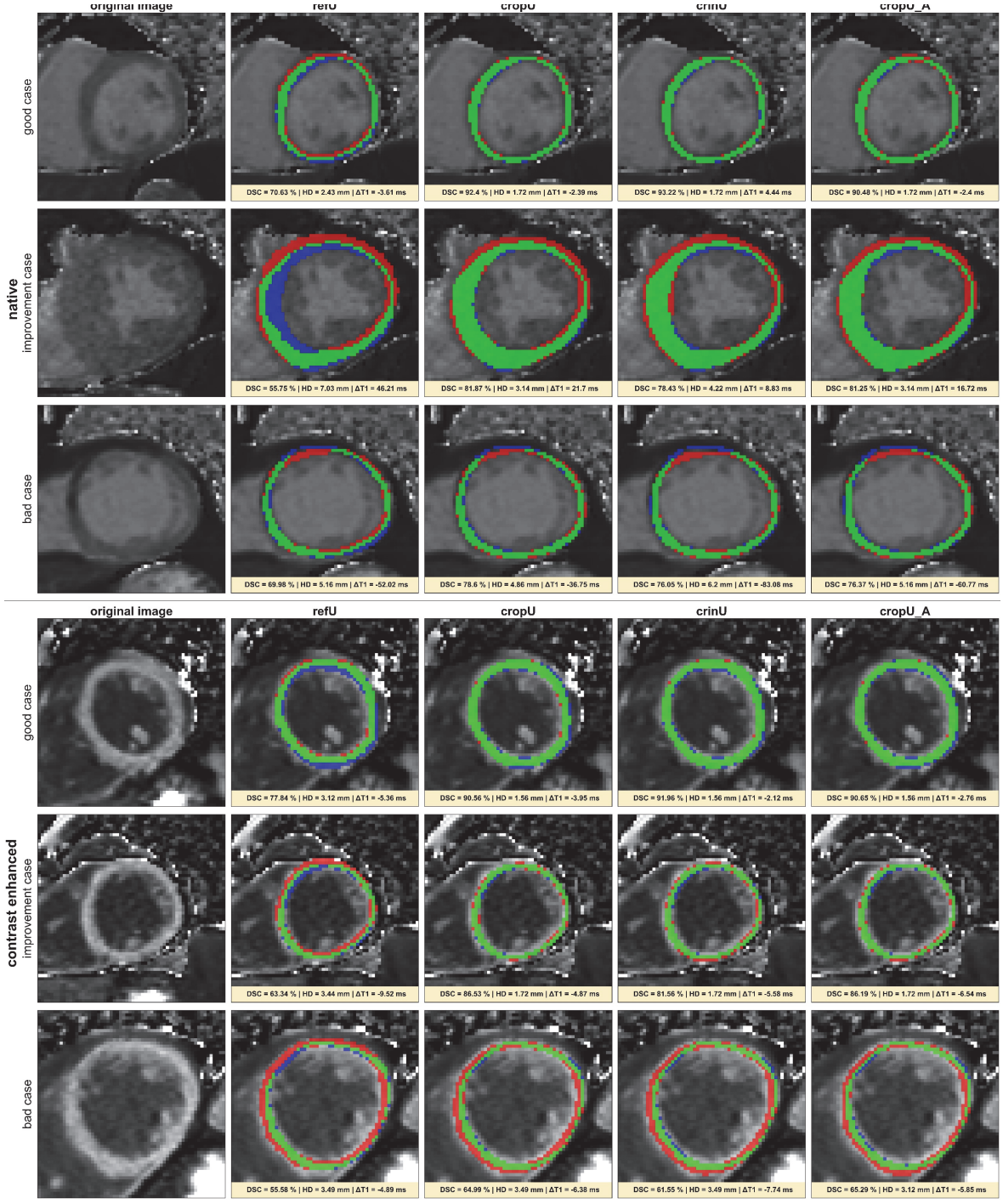

Figure 8: CASEG results for refU, cropU, crinU and cropU_A as exemplary plots

A good, a bad and an improving case for native and contrast enhanced T1 maps are respectively displayed. Green denotes consensus with expert segmentation, red denotes false positive and blue denotes false negative segmentation. Figure adapted from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0

The overall geometric quality improved significantly in the DSC from around 72% towards 80% and the HD reduced significantly from around 3.70mm to less than 3.35mm for all CASEG pipelines compared to refU. The boxplots in Figure 9 show the value distribution of achieved DSC and HD values. All CASEG pipelines reached in rare cases a DSC beyond 90% and consequently exceeded the maximum DSC of 86.39% in native and 83.56% in contrast enhanced T1 maps by refU. On the contrary, a DSC of 70% or higher is conventionally assumed as good [46], which is not reached in all cases across the CASEG pipelines. Although the HD reduced significantly, it remained on average in a local segmentation deviation of two to three voxels.

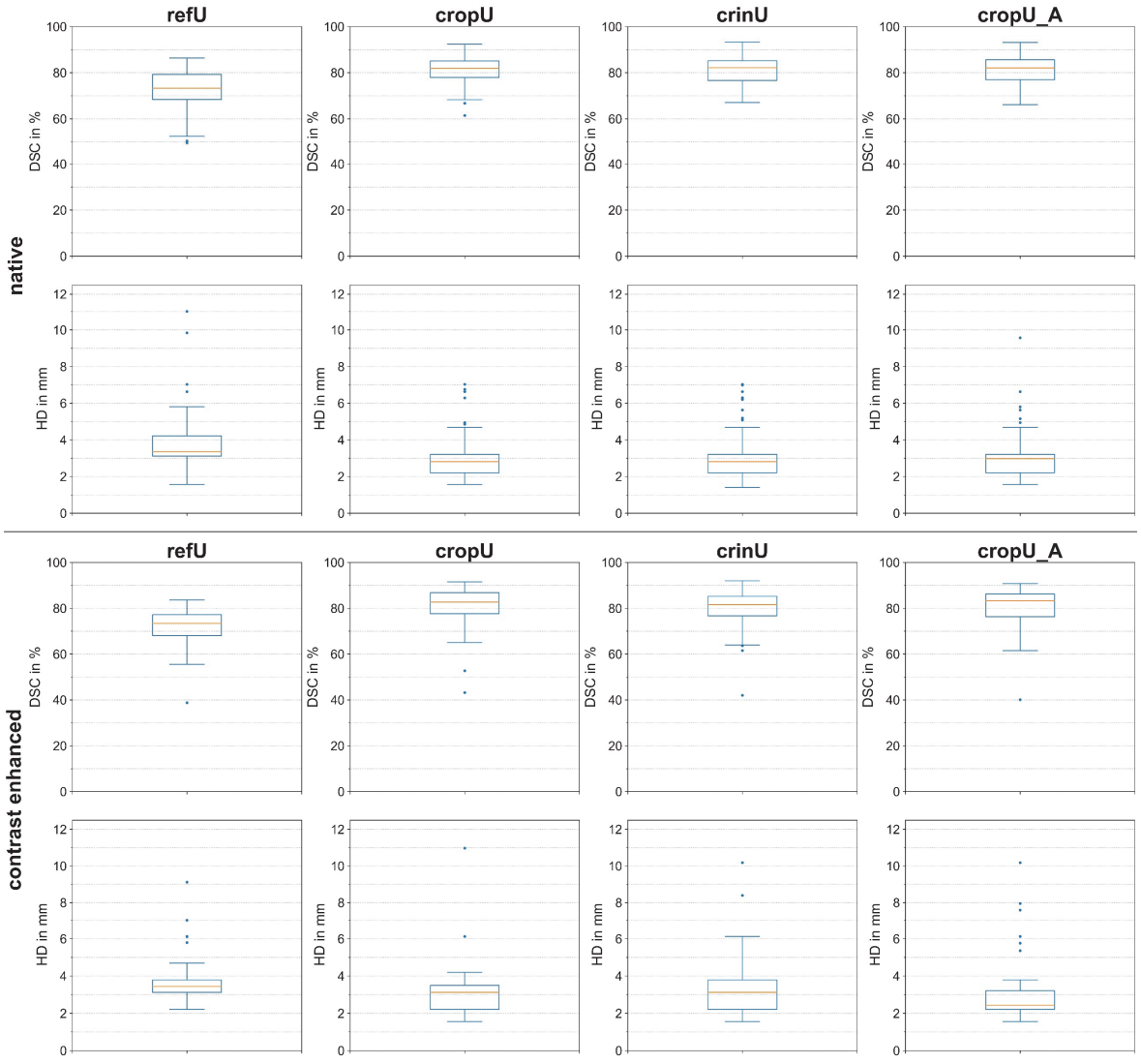

Figure 9: CASEG results for refU, cropU, crinU and cropU_A in the geometric domain

Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) were separated for native and contrast enhanced T1 maps. Figure adapted from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0

The quantitative metrics ME, MAE and RMSE were reduced in the CASEG pipelines compared to refU except for the MAE on contrast enhanced test data in crinU. Apart from the ME in the native test data on crinU and cropU_A, ME and MAE were not significantly improved. The CI remained in the accepted tolerance range of 24.5ms[43]. As this equivalence margin was defined for native data only, an adhered equivalence margin for contrast enhanced T1 maps was assumed due to a lowered COV compared to the native T1 maps. Nonetheless, the correlation and Bland-Altman plots in Figure 10 show among the native cases 20 in refU, 11 in cropU, 12 in crinU and 11 in cropU_A and 1 contrast enhanced case among all pipelines that exceeded the 24.5ms limits of equivalence.

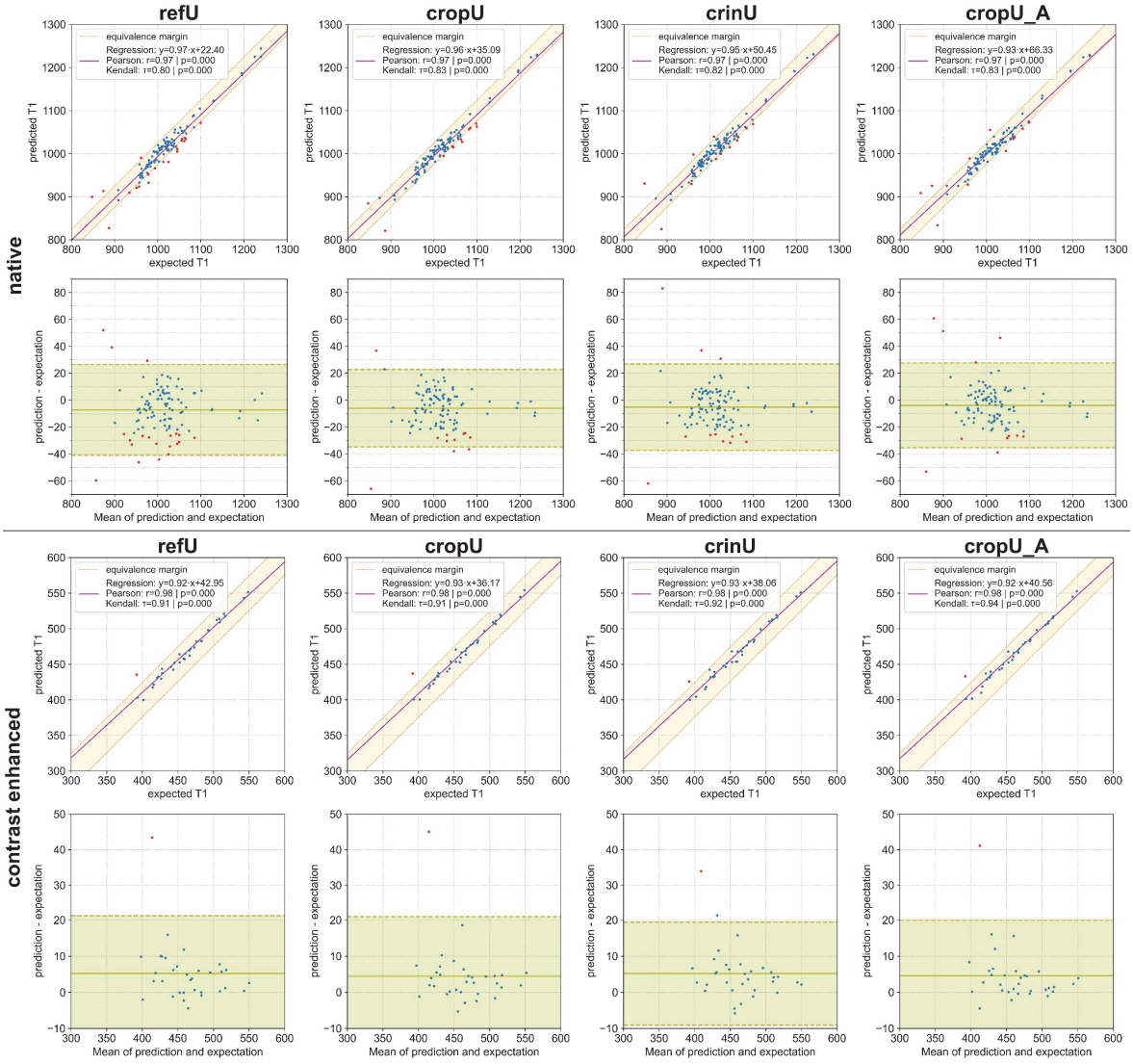

Figure 10: CASEG results for refU, cropU, crinU and cropU_A in the quantitative domain

A correlation and a Bland-Altman plot separated for native and contrast enhanced T1 maps is displayed. Blue denotes datasets within the accepted tolerance ranges and red denotes exceeding them. Figure adapted from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0

All segmentations showed with significance a very strong linear correlation, in native T1 maps a strong and in contrast enhanced T1 maps a very strong monotonic correlation. A COD of at least 94.09% explains the majority of the predicted average T1 value variation by the intrinsic LV myocardial T1 time. The limits of agreement in the Bland-Altman-plots showed only slight differences between the different segmentation models. The equivalence margin exceeding native cases were mostly underestimating the expected average T1 value and aligns with the histogram of disjoint segmented pixels in Figure 11.

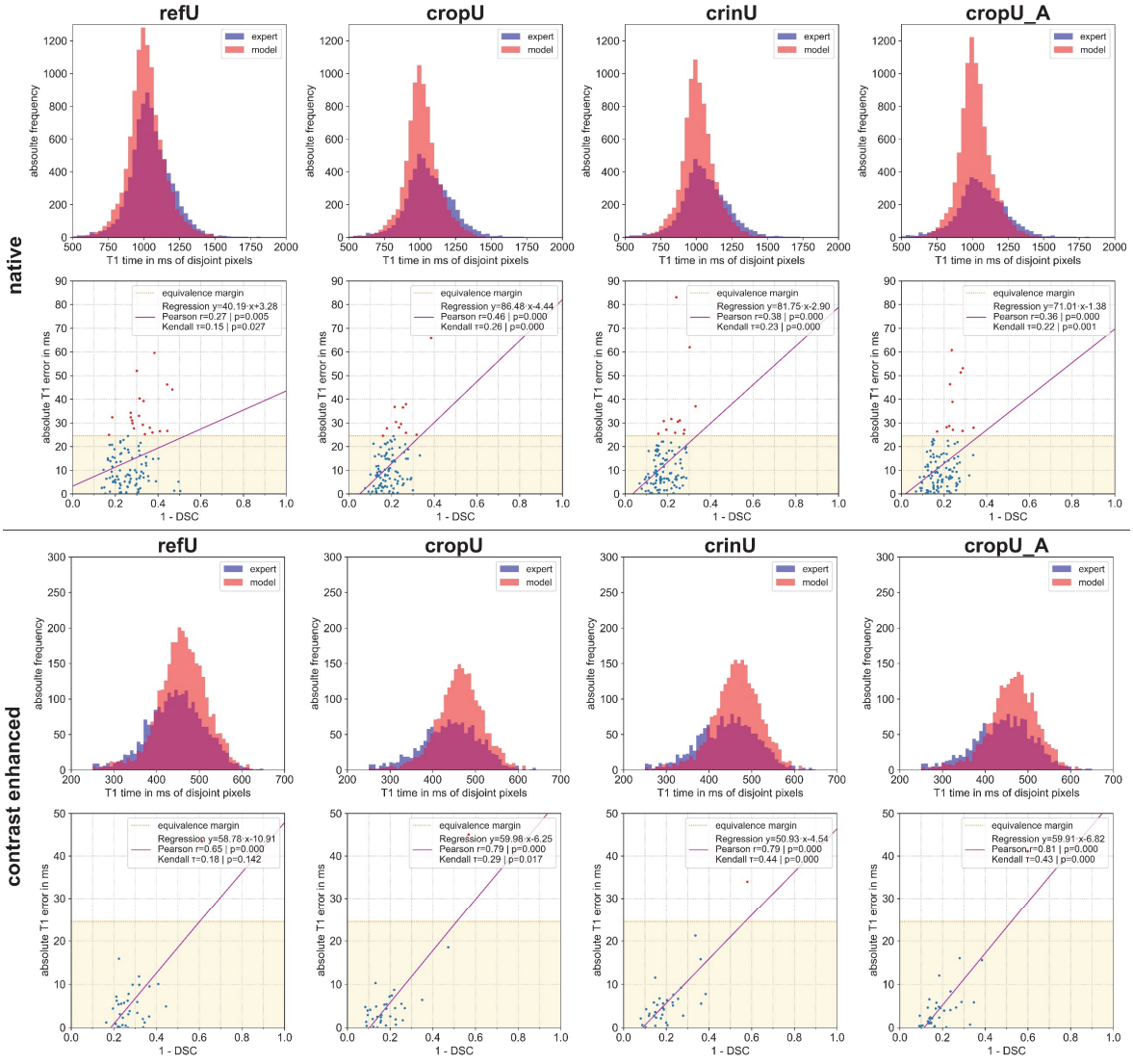

Figure 11: CASEG results for refU, cropU, crinU and cropU_A as relationships

The histograms show the disjoint segmented pixels clipped at 2000ms while some rare individual pixels exceeded that margin. The correlation plots reveal the relationship between the geometric and quantitative domain, represented by the Dice Similarity Coefficient (DSC) and the mean absolute error (MAE). The blue dots represent cases within the accepted tolerance ranges while the red dots exceeding them. Figure adapted from "Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance" by Viezzer et al.[54], 2023, under CC BY 4.0

The additional coherence plots in Figure 11 depict the relationship between the geometric and quantitative domain, represented by the DSC and MAE. The linear correlation between both metrics was weak and moderate in refU while moderate and strong in all CASEG pipelines for native and contrast enhanced T1 maps respectively. The monotony correlation was weak for all models except for a moderate correlation of contrast enhanced data in crinU and cropU_A. All correlation coefficients, except for the monotony correlation of the contrast enhanced data in refU, were statistically significant. A maximum Pearson correlation of 0.81 implied a COD of 65.61% revealing an inexplicability of more than a third of the MAE variation by the DSC. Finally, the majority of the 24.5ms equivalence margin exceeding cases had a DSC of more than 70%, which was assumed with a good geometric concordance[46].

3.2 Cascaded Segmentation (CASEG) in the Magnetic Resonance Imaging Software for Standardization (MARISSA)

The used datasets for MARISSA were segmented with a retrained cropU pipeline. The retraining included the original training and testing data together and was validated on the validation data resulting in a DSC of 83.01% and 82.20% for the novel training and validation data after 156 epochs.

Among the M=970 Healthy T1 maps in MARISSA 180 segmentations from the retrained cropU segmentation, 90 segmentations from the research segmentation model by Siemens Healthcare, 665 segmentations of common intersected segmentation among the retrained cropU and research segmentation and 35 manual segmentations by experts were used. Regarding the M=145 T1 maps from patients with HCM and AMY 85 segmentations from the retrained cropU, 4 segmentations from the research segmentation model by Siemens Healthineers, 35 segmentations of common intersected segmentation among the retrained cropU and research segmentation and 21 manual segmentations by expert were used.

In conclusion, the retrained cropU CASEG segmentations were useful in 965 out of 1115 cases (86.54%) and in combination with a separate segmentation model only 5.02% of the cases required a manual segmentation intervention. Nonetheless, a visual inspection was still required.

3.3 Magnetic Resonance Imaging Software for Standardization (MARISSA)

The implementation of MARISSA in Python with a PyQt5 based GUI and a SQLite database backend was completed and tested for general functionality as well as its im- and export features. The following results on post-hoc standardization of native parametric T1 maps are retrieved from the publication by Viezzer et al.[52].

3.3.1 Training in MARISSA

The 24 standardization pipelines without clustering were successfully trained. However, the individual mode could not consider the T1 map MOLLI 3(3)5b and T1 map SASHA GRE sequences due to mismatching variations in the other CPs among the HTR during training. As a consequence, the individual mode pipelines were not able to standardize for those two sequences. The top three performing standardization pipelines were LSVR regression on relative values in cascaded mode, ETR on relative values in ensemble mode and ETR on absolute values in ensemble mode with a COV reduction of the average T1 values in HTE from 12.47% towards 5.98%, 6.10% and 6.23% respectively.
The BPSP among the 240 trained pipelines were, finally, the LSVR regression on relative values in the cascaded mode with two bins and the agglomerative single clustering resulting in a further reduced COV by 0.17% towards 5.81% in the HTE. Figure 12 illustrates all HTE based COVs of the 240 trained standardization pipelines and reveals that the standardization pipeline setting is crucial for the outcome. Some standardization pipeline settings even worsened the intra-cohort variation of HTE especially in combination with clustering.

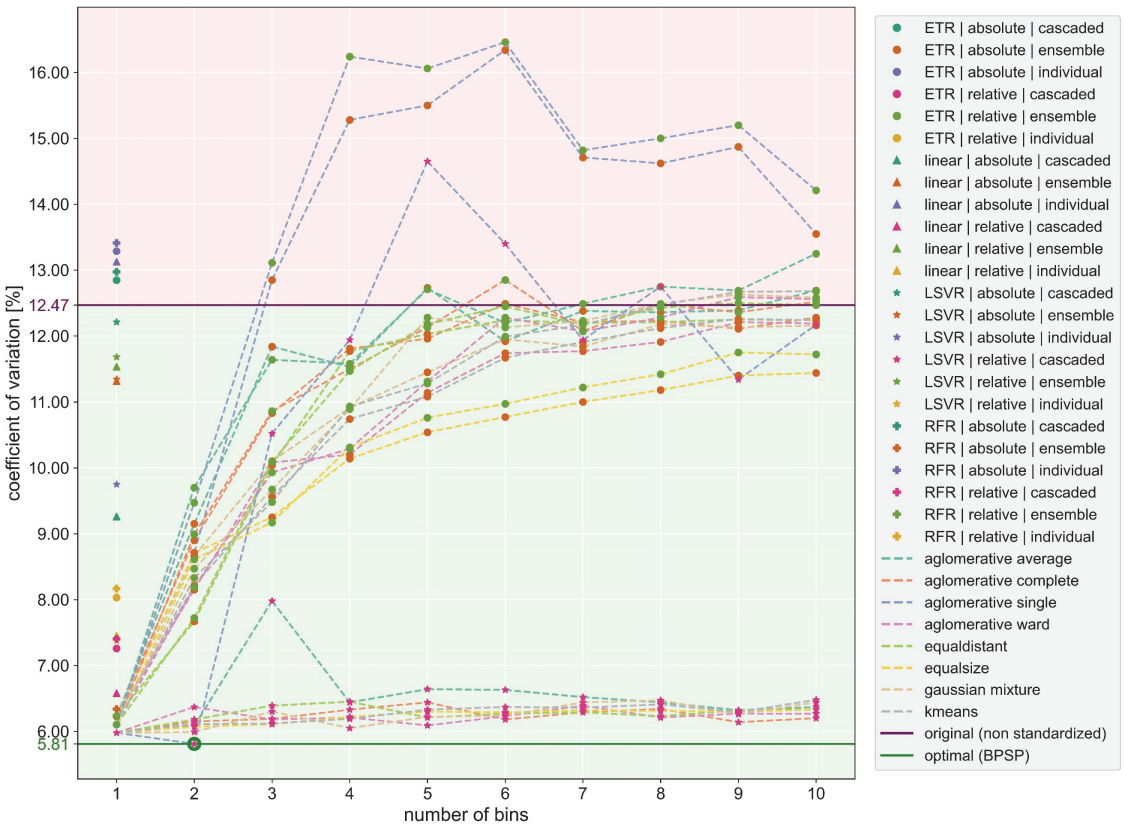

Figure 12: MARISSA results on the 240 trained standardization pipelines

The coefficient of variation (COV) on the healthy test dataset is displayed for various standardization pipeline settings. The light green area denotes an improvement and the light red area a worsening compared to the original data. Only the top three performing settings were tested for different clustering results on up to 10 clusters. The original (purple) and best (green) COV values are highlighted. Figure retrieved from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

3.3.2 Testing of MARISSA

By application of the BPSP on the HCM and AMY cohorts, the COV were reduced from 9.56% and 6.06% towards 4.46% and 6.05% respectively. Consequently, the COV reduced in the HTE and HCM cohorts but remained almost equal in the AMY cohort.

While HTE and HCM as well as HCM and AMY but not HTE and AMY were already statistically significant different before standardization, the standardization with the BPSP resulted in a statistically significant difference among all cohort combinations. Furthermore, the 25% to 75% quantile ranges within HCM (999.57-1186.00ms) and AMY (1087.84-1148.40ms) overlapped for the most part with HTE (1007.81-1213.34ms) before standardization; although, higher T1 values were expected in the patient groups[105]. This dichotomy origins in the variations of the CPs within the cohorts especially regarding the sequence and system and explains why current guidelines recommend local reference values[20]. After application of the BPSP, the 25% to 75% quantile ranges reached

1112.02-1167.72ms in HTE, 1157.81-1226.76ms in HCM and 1308.96-1394.79ms in AMY reflecting the expected distribution. The distributions including the CIs and the CP impact for each case are shown in
Figure 13. Although all CIs exceeded the CI of the unstandardized HTR data whose CPs matched the reference CP environment, the HTE CI after standardization was most likely to fit while HCM and AMY data were clearly above. Nonetheless, the average T1 value ranges after standardization of 1136.78±66.09ms, 1186.27±52.93ms and 1337.62±80.92ms for HTE, HCM and AMY respectively as well as the boxplots in
Figure 13 reveal overlapping value ranges.

Consequently, the ROC analysis, as shown in Figure 14, identified a threshold (sensitivity/specificity) of 1163.89ms (71.90%/72.44%) between HTE and HCM, 1204.46ms (95.83%/91.67%) between HTE and AMY and 1287.89ms (87.50%/98.35%) between HCM and AMY. The 150% margin for the sum of sensitivity and specificity to reach a good diagnostic performance[97] was obtained for the diagnostics between HTE and AMY as well as HCM and AMY but slightly missed by 5.66% between HTE and HCM. However, in all cohorts, the sensitivity and specificity among all scanner and sequence variants captured values in the range of those from unstandardized intra-scanner-intra-sequence data. Additionally, among the intra-scanner-intra-sequence datasets, the sum of sensitivity and specificity almost remained or even improved while either value changed.

Eight subjects of the HTE group were examined under two or more different technical conditions. Figure 15 shows for each of the eight subjects the original examined T1 values and their respective standardized values as well as the T1 values spread and the intra-subject COV. Except for one case, all T1 value spreads could be minimized, which is also reflected in a COV minimization across all subjects. However, individual measurements were outlying after standardization and mostly concern SASHA based T1 mapping sequences reflecting its known lack of precision in favour of the accuracy[33].

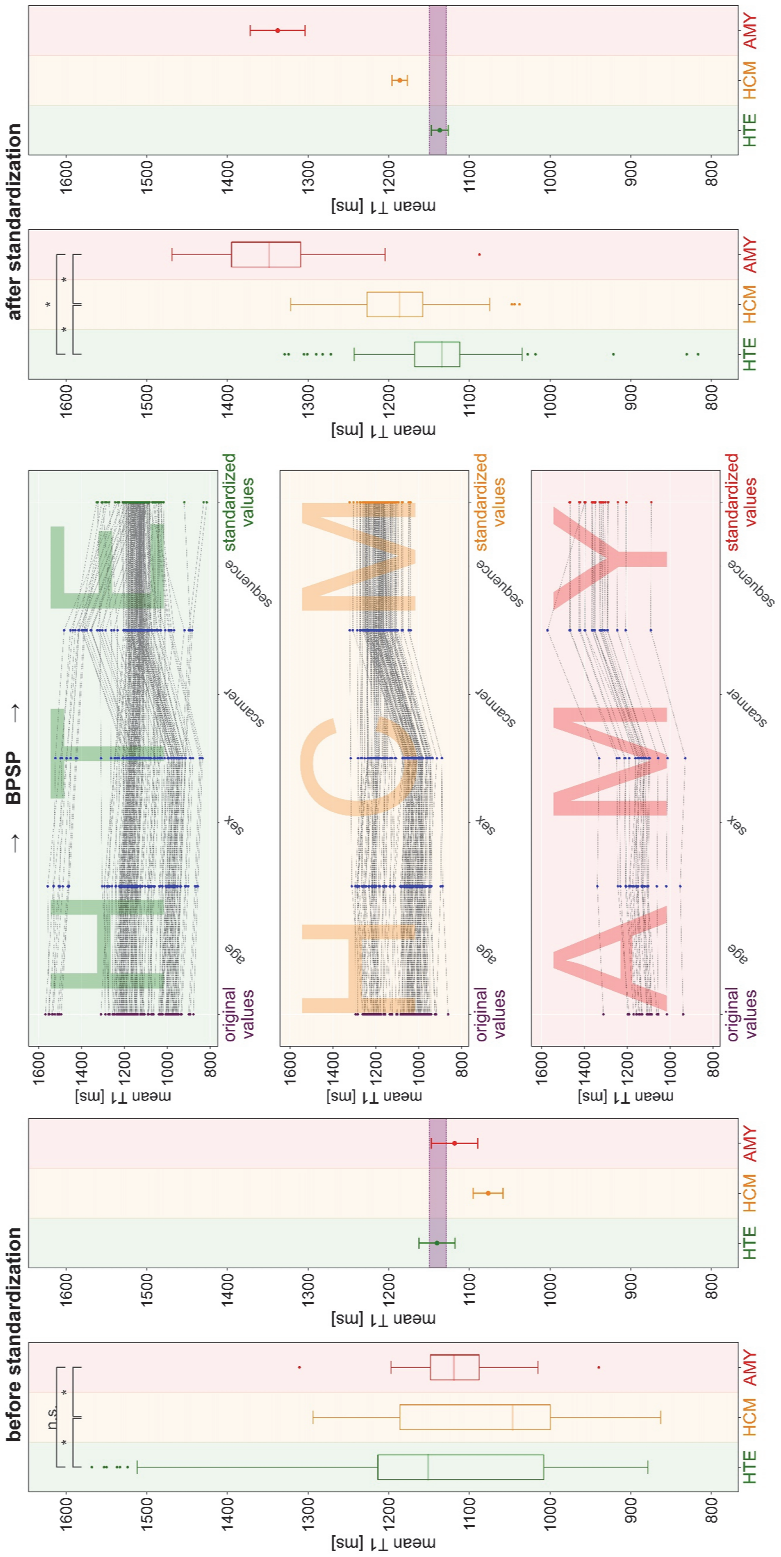

Figure 13: MARISSA results for the best performing standardization pipeline (BPSP)

*The boxplots and confidence intervals are given for before and after standardization with a central plot showing the propagation of the T1 values with respect to each confounding parameter on the respective cohort healthy (HTE), patients with hypertrophic cardiomyopathy (HCM) and patients with amyloidosis (AMY). * denotes statistically significance, n.s.: non significant. Figure retrieved from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.*

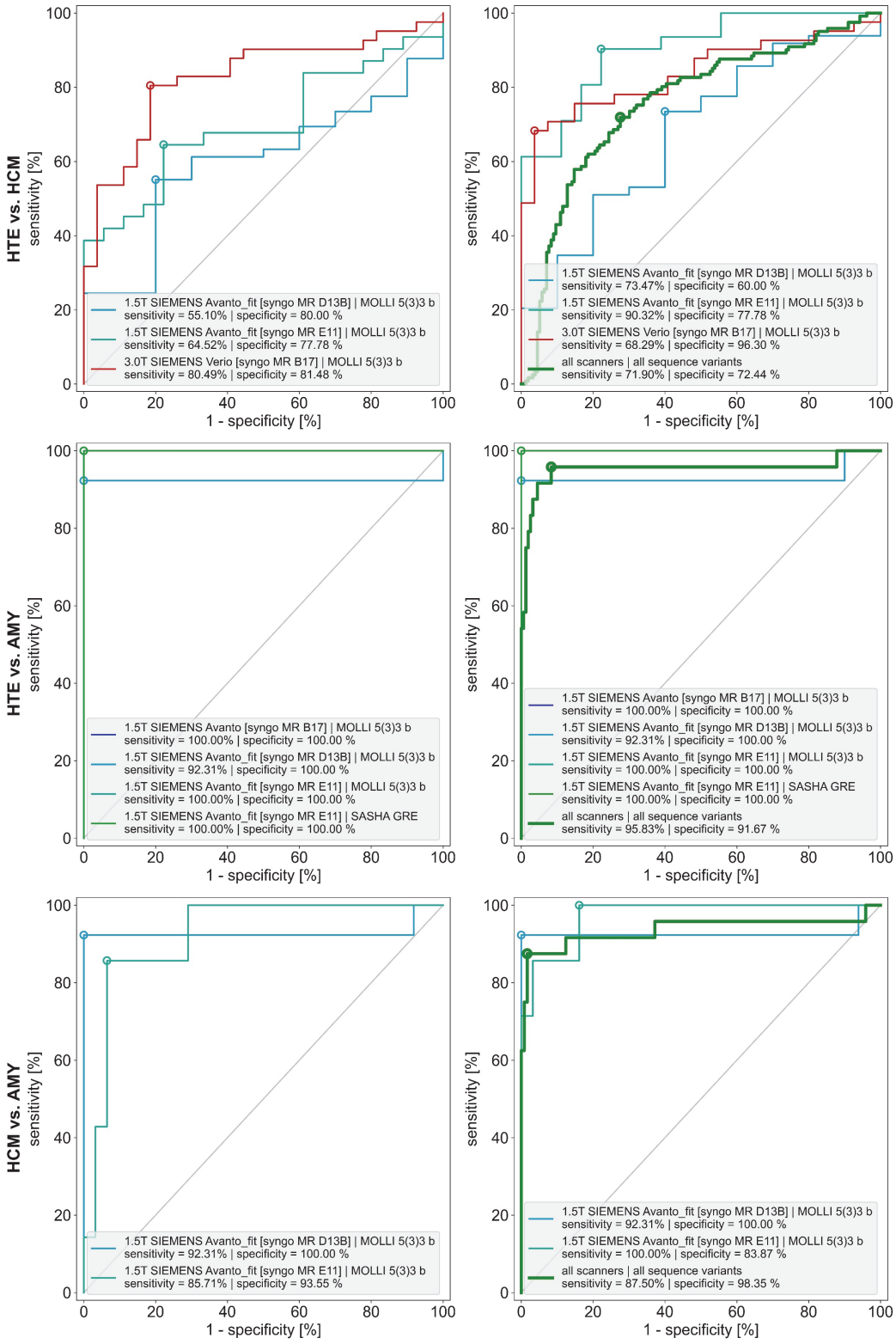

Figure 14: MARISSA results for the best performing standardization pipeline (BPSP) as receiver operating characteristic (ROC) analysis

The intra-scanner-intra-sequence plots for before and after standardization were complemented with the ROC analysis among all data after standardization considering the healthy (HTE), patients with hypertrophic cardiomyopathy (HCM) and patients with amyloidosis (AMY) cohorts. Image retrieved from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

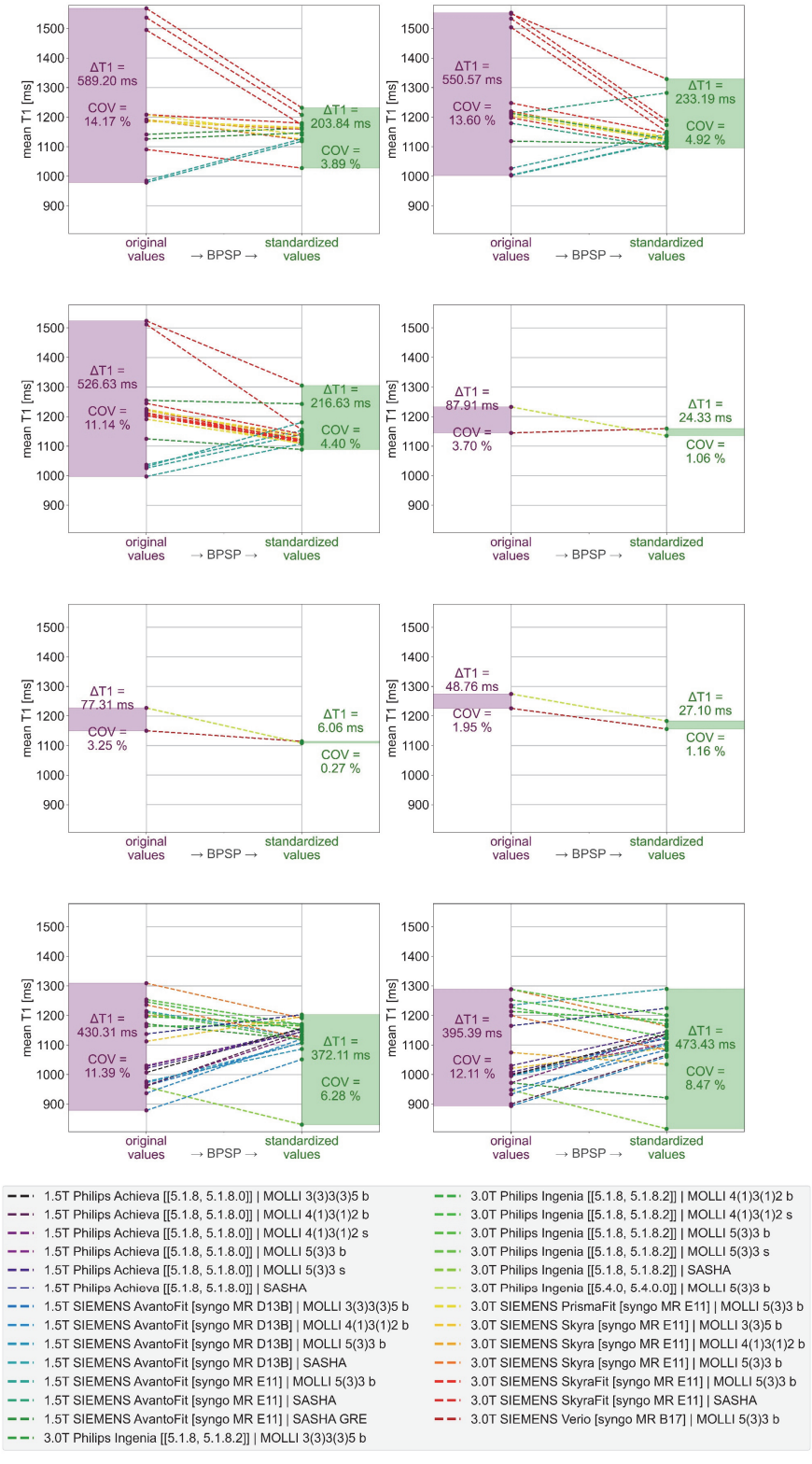

Figure 15: MARISSA results for the best performing standardization pipeline (BPSP) among the healthy (HTE) subjects with multiple, different acquisitions

The value spread and coefficient of variation (COV) are given for before and after standardization. Figure retrieved from "Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept" by Viezzer et al.[52], 2024, under CC BY 4.0.

4 Discussion

This work aimed to minimize the CPs induced biases in parametric T1 mapping by the automatic segmentation approach CASEG and a regression model based post-hoc standardization with the MARISSA.

An improved performance of automatic segmentation approaches is achieved by ever-more complex network structures[46,47,118] or increased input data quality[119–121]. The CASEG incorporated the latter by an upstream image section focusing ODA with a significant improvement of the geometric performance compared to a reference segmentation model. The quantitative domain improved likewise, however, not statistically significant and individual cases still exceeded accepted tolerance ranges. Consequently, a full substitution of manual segmentation by CNNs requires further investigation, especially in the light of atypical errors compared to human readers[56].

A retrained CASEG cropU model was used in combination with an independent, alternative fully automated segmentation model to generate the segmentations of the considered MARISSA datasets. A visual inspection was necessary to approve the segmentation quality and in rare cases manual corrections were performed by an expert.

As a proof-of-concept, MARISSA was used to harmonize myocardial native T1 values of midventricular slices in CMR with respect to the CPs age, sex, scanner and sequence. The evaluated post-hoc standardization pipelines in MARISSA identified the pipeline settings as crucial for the outcome quality. The evaluated BPSP was set up as LSVR on relative values in the cascaded mode with two bins by an agglomerative single clustering. After application of the BPSP, the COV was halved in the HTE cohort while the intra-scanner-intra-sequence diagnostics among HTE, HCM and AMY almost remained or even increased. Consequently, the feasibility of a post-hoc standardization of parametric T1 maps in CMR was shown.

In the following, the results of the CASEG approach are discussed first, followed by an illumination of the BPSP outcomes in MARISSA, the limitations of both and finally an outlook on their potential future integration in the CMR imaging chain is shown.

4.1 Cascaded Segmentation (CASEG)

The manual and time-consuming segmentation of parametric T1 maps suffers from intra-reader variabilities[46,47]. In recent years, CNNs were increasingly used to automate this task and stirred up hope for accurate and reproducible segmentations[60,122]. However, uncertainties persists due to deviations from expert segmentation and differences across CNN models[123]. The CASEG approach maintained the precision advantages of a CNN while simultaneously increasing the accuracy by focusing on the relevant image section. The novelty of the CASEG approach is the integration of a preliminary ODA for the parametric T1 map segmentation. The BB detection by an ODA have been widely used in other disciplines considering a plethora of objects within an image resulting in highly complex network structures[124,125]. Due to the single detection of the LV as the only object class, the simple, classical U-Net structure was sufficient. The geometric concordance is in line with the DSC results of 92.4±3.6% for the LV BB detection in native CINE images by Niu et al.[126]. The false-negative uncertainty was reduced by the magnification factor enlargement, which is optimized for the used dataset and does not necessarily generalize for other data. While too low and too high magnification factors impar the overall performance, cropU and crinU showed a robust performance in a factor range of 1.3 to 2.5.
Although CASEG could be implemented as a fully complex end-to-end network structure like the DoubleU-Net[127], the separation of ODA and segmentation network enables an easy replacement of either network but requires intermediate processing steps.
The classical U-Net, like the refU, is a common automatic segmentation CNN for medical images and was used in literature as a benchmark for novel complex CNNs[46–48]. Although a similar structure was used among all those studies, the hyperparameter settings were different. The evaluation of optimal hyperparameters is a tedious task that was systematically approached in this work. Consideration of unaddressed hyperparameters requires more effective approaches as described in literature[128–130] to handle the exponential increasing amount of hyperparameter combinations. Additionally, the training of the U-Nets based on a pure geometric loss function which misappropriates the quantitative domain.
In the test dataset only midventricular and basal slices were used as those were recommended by current guidelines as stable slice location[20]. In literature, alternative automatic segmentation models used either midventricular slices only[47] or the whole SAX stack[46,48], but in any way different datasets that were based on a ShMOLLI acquisition scheme. Although ShMOLLI is a MOLLI derivative, the shortened version produces different T1

values than other MOLLI schemes[32] and thus limits the quantitative comparison with those alternative segmentation models.

In native parametric T1 maps the refU performed inferior compared to the U-Net DSC results of 78%[47], 82.7%[48] and 83.13%[46] according to literature. However, the incorporation of the BB information by cropU, crinU and cropU_A pipelines improved the geometric results compared to refU and aligned with the performances of the standard U-Nets in literature[46–48]. The same applies to contrast enhanced T1 maps with refU being inferior to the DSC of 74.1% by Farrag et al.[48] while the CASEG pipelines including the BB information were outperforming. Nonetheless, the DSC results of 84% by Puyol-Antòn et al.[47] and 85% by Hann et al.[46] with their respective complex network structures were out of reach for the CASEG pipelines. Compared to a human reader, the CASEG pipelines showed comparable results to the intra-observer performance of 72% DSC and 15.61mm HD in native and 83% DSC and 9.03mm HD in contrast enhanced data[131].

Although the CASEG pipelines showed an improved geometric performance compared to refU, the quantitative performance was not significantly improved. Hann et al.[46] reached an even higher DSC of 85%, but the MAE of 11.3ms were only slightly below the cropU results. Considering the maximal reached COD of 62.41% between DSC and MAE, a systematic minimization of quantitative outcomes is only achievable at higher geometric concordance. The MEs in the CASEG pipelines as well as refU were in the published range of 4.6ms[46], 8ms[48] and 12.4ms[47] for native T1 maps but exceeded the 2ms[48] in contrast enhanced data while the CIs remained in the intra-observer equivalence range[43]. The used equivalence margin was the strictest applicable rule as it reflected an expert intra-observer variability[43]. It was, however, solely evaluated for native T1 maps. The application on contrast enhanced T1 maps in combination with a COV comparison is a similitude but not scientifically approved.

The annulus shaped segmentation ROI accounts for all LV segments at the expense of increasing variation compared to a septal segmentation[20]. False positive segmented voxels at either epi- or endocardial border potentially further impair quantified T1 values due to partial volume effects. Commercial software, like cvi42 (Circle Cardiovascular Imaging, Calgary, Canada), provides a safety margin functionality to shrink the segmented contours towards intra-myocardium by a defined proportion to compensate this effect on the potential cost of turning some true positives into false negatives. This post-processing step has not been addressed in the CASEG pipelines so far.

4.2 Magnetic Resonance Imaging Software for Standardization (MARISSA)

While the comparability of parametric T1 mapping data was published over the last decade, the validity was only limited to specific cohorts and technical environments[40,132–134]. The z-score transformation into values without a unit of measure was a first generalizable approach for a CP independent T1 value comparability[51]. Although this transformation is considered in current guidelines[20], its establishment is undermined in the clinical routine due to additional efforts and costs. A local healthy cohort examination is necessary to evaluate the reference values for the z-score transformation[51] and potentially requires a re-examination after a hard- or software change. Similar to the fixation of the technical environment, the z-score does not account for subject specific CPs. Physiological CP influences on parametric T1 mapping has been described in literature by regression model analysis[39,40]. At this point MARISSA comes into play and harnesses regression models for a post-hoc standardization. This approach also relies on healthy volunteer data but reveals a generalization potential. Technical as well as subject specific CPs were considered and their assessed influences are transferable onto other sites. This potentially reduces the amount and costs in a long-term perspective.

The post-hoc standardization in MARISSA was limited to the four considered CPs age, sex, scanner and sequence. As processing and reconstruction performed on the scanner are usually not described in DICOM tags, those potential CPs are currently neither includable nor accessible. The segmentation procedure, in turn, is accessible but hard to define. The segmentation was circumvented as a CP by an automatic segmentation with sparse manual interventions. Nonetheless, other CPs, like heart rate (HR)[33,135], body-mass-index (BMI)[136] or partial volume effects due to voxel sizes[137], are known to influence parametric T1 mapping values. Due to anonymization, BMI and HR were partly missed in the considered datasets. As these three CPs are numerical parameters, additional datasets would be necessary to capture all categorical CP values in the individual and cascaded mode. The amount of necessary training data in MARISSA depends in particular on the number of considered CPs, the value variation in categorical CPs and the mode of the standardization pipeline. The individual mode accounts best for isolated CP biases, but requires the highest amount of training data with well sorted CPs variations in order to capture all CP values. The two missed sequences revealed a lack of training data for the evaluated individual mode pipelines. The cascaded mode is more robust than the individual mode and potentially depicts inter-CP dependencies in parts but may miss CP values

due to its susceptibility to the order of considered CPs. The multi-CP regression model in the ensemble mode requires the least number of training data and has the best potential to depict inter-CP dependencies. However, the ensemble mode standardization fails for data that includes categorical CPs of unknown value while cascaded and individual mode are able to skip for those and standardize in regard of the remaining CPs. The conversion of numerical CPs into categorical by clustering i.e. ages into decades or BMI into groups proposed by the world health organization[138] may in turn decrease the necessary amount of training data.

In addition to the mode, the choice of the regression-type and unit of measure were most important for improving the COV by 6.49%, whereas the value clustering reached an additional gain of only 0.17%. Even worse, clustering was shown to be detrimental on the performance in many cases, which shows that this is a rather potential fine-tuning step than a necessary option for the post-hoc standardization.

The used datasets included only Philips and Siemens scanners as other manufacturers were not accessible. The patient cohorts, HCM and AMY, were skewed since examinations were mainly performed on 1.5T Siemens scanners and thus had lower T1 values than expected compared to healthy subjects[105].

After application of the BPSP, the expected statistically significant differentiation across the considered cohorts HTE, HCM and AMY with respective increasing T1 values[105] were reached. The ROC analysis results were comparable to intra-scanner-intra-sequence circumstances, but the 150% margin for sufficient evident differentiation[97] was narrowly missed between HTE and HCM. As hypertrophic cardiomyopathy rarely affects the whole LV, the focus on midventricular slices only does not reflect the plethora of phenotypes of HCM[139]. Additionally, a variety of genotypes and risk factors affects the state of a hypertrophic cardiomyopathy and consequently, the affected myocardium changes over time with manifold morphologies[106,107]. For that reason, the partly overlapping but statistically significant different T1 value ranges of HTE and HCM align after standardization with intra-scanner literature values on 3T scanners (Healthy vs. HCM) by Liang et al. (1228.4±42.7ms vs. 1290.0±64.3ms)[103], Qin et al. (1240.0±29.8ms vs. 1308.0±55.5ms)[104] and Lavall et al. (1225±21ms vs. 1266±44ms)[105].

The differentiation between HTE and AMY after BPSP standardization was in the range of Baggiano et al.[140], with an intra-scanner sensitivity of 85% and specificity of 87%, and Kranzusch et al.[51], with a sensitivity of 96% and specificity of 100% after application of

the z-score. However, Baggiano et al.[140] included 436 patients with amyloidosis compared to solely 38 in the work by Kranzusch et al.[51] and 24 in this work.
The sensitivity of 76.1%, 86.54% and 100% as well as the specificity of 80.36%, 83.3% and 97% by Lavall et al.[105], Nam et al.[141] and Martines-Naharro et al.[142] suits the respective obtained values between HCM and AMY in this work. Consequently, after standardization of midventricular native parametric T1 maps with the BPSP patients with amyloidosis are reliable diagnosable while hypertrophic cardiomyopathy detection depends on the amount and location of fibrosis.
The persistence of outliers after standardization in an intra-subject perspective reveals imprecision as a major limit. Therefore, optimal results are solely obtained with precise data. This requires a respective sequence variant on a stable, high quality scanning system. This susceptibility to imprecision is shared by any post-hoc transformation approach, like the z-score[143], as long as no further CP information is available that correlates with the source of imprecision.

4.3 Limitations

A selection procedure for the used datasets in regard of native and contrast enhanced T1 maps in CASEG and CP value distributions as well as cohort statistics in MARISSA was omitted in favour of using as much available data as possible. Consequently, the datasets were imbalanced with respect to cohort characteristics, scanner and sequence variants. This limits the generalisability of the CASEG models as well as the BPSP results. Although scanner-sequence combinations that have not been reflected in the training data would pass through the BPSP with a standardized output, a dedicated investigation remains unaddressed yet. The evaluated output was limited to midventricular slices only in case of MARISSA and midventricular and basal slices in case of CASEG.
The BPSP transforms quantitative parametric T1 mapping values into values of a reference CP environment acquisition without consideration for accuracy. The systematic bias between the reference CP environment and the intrinsic true T1 value of the underlying tissue persists. The reference value for the sequence, for example, was set to MOLLI 5(3)3 b, which is commonly known as a precise but underestimating sequence[33].

4.4 Outlook

Parametric T1 mapping lacks universal valid reference values due to CP influences along all steps in the imaging chain[20,22]. A harmonization of those minimizes the induced biases. While assessment specific CPs were handled by the CASEG approach, the BPSP within MARISSA accounted for subject-specific and technical CPs. If in a long-term future perspective both, the fully automated segmentation procedure and the post-hoc standardization approach, show reliable and validated results without the necessity of a manual intervention, then the CMR imaging chain will be adapted. The manual segmentation task will be substituted by an automatic approach while a standardization step will be inserted between the segmentation and classification. Preferably, both will be implemented inline the MRI scanner and provided together with the source images as well as the original T1 map. Heretofore, further optimizations are worth investigating.

Regarding the CASEG, performance gains are reachable by replacing the secondary segmentation CNN with more complex architectures[46,47] or recent U-Net model adaptions[118]. The crucial border pixel segmentation may be improved by splitting the secondary segmentation into two subtasks in order to segment the endocardial contour, represented by the blood pool area, and the epicardial contour, represented by the blood pool and myocardial area, independently. Finally, an essential step for the automatic segmentation of parametric T1 maps is the definition of a loss function that includes the geometric and quantitative domain at the same time. As the absolute values differ due to the different unit of measures, the definition of such a loss function is not straight forward.

In MARISSA automatic segmentation strategies must be considered as a potential CP in the future since artificial intelligence based segmentations become more popular[144], especially in the view of increasingly used datasets. The results of the BPSP were only a first step to prove the concept and feasibility of a post-hoc standardization. In the future, more training data must be included in order to access more CPs and other CP values. A validation on datasets of similar scanners and sequences from other sites are necessary to prove the generalizability among scanners of the same vendor. MARISSA has the potential to strengthen the use of parametric T1 mapping in the clinical routine by making T1 mapping values transferable. Finally, as the name Magnetic Resonance Imaging Software for Standardization (MARISSA) implies, the post-hoc standardization approach is implemented for the usage on other quantitative methods like parametric T2, T2* or T1ρ mapping, but requires further investigation.

5 Conclusions

Segmentation is a necessary step for the evaluation of parametric T1 maps. While fully automated segmentation procedures play an increasing role to accelerating and simplifying the workflow, the usage of a bounding box information in a segmentation model, like the cascaded segmentation (CASEG), improves the segmentation quality. The quantitative domain is underrepresented during training and requires further development of suitable loss functions. A visual inspection is currently inevitable to assure quality as done for the used datasets in the regression model based post-hoc standardization approach. The standardization pipeline performance highly depended on an appropriate setting. The evaluation of the best performing standardization revealed the comparability of datasets of different origin while maintaining the diagnostic power as in intra-scanner-intra-sequence circumstances among healthy subjects and patients. Further improvements are expected in the future by including more datasets and confounding parameters in the standardization pipeline training of the Magnetic Resonance Imaging Software for Standardization (MARISSA).

Reference list

1. World Heart Federation. *World Heart Report 2023: Confronting the World's Number One Killer.* (World Heart Federation, 2023). at <https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf>
2. Heidenreich, P. A., Bozkurt, B., Aguilar, D., Allen, L. A., Byun, J. J., Colvin, M. M., Deswal, A., Drazner, M. H., Dunlay, S. M., Evers, L. R., Fang, J. C., Fedson, S. E., Fonarow, G. C., Hayek, S. S., Hernandez, A. F., Khazanie, P., Kittleson, M. M., Lee, C. S., Link, M. S., Milano, C. A., Nnacheta, L. C., Sandhu, A. T., Stevenson, L. W., Vardeny, O., Vest, A. R. & Yancy, C. W. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation* **145,** e895–e1032 (2022).
3. McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., Cleland, J. G. F., Crespo-Leiro, M. G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak, M., Lam, C. S. P., Lyon, A. R., McMurray, J. J. V., Mebazaa, A., Mindham, R., Muneretto, C., Francesco Piepoli, M., Price, S., Rosano, G. M. C., Ruschitzka, F., Skibelund, A. K., & ESC Scientific Document Group. 2023 Focused Update of the 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. *European Heart Journal* **44,** 3627–3639 (2023).
4. Arbelo, E., Protonotarios, A., Gimeno, J. R., Arbustini, E., Barriales-Villa, R., Basso, C., Bezzina, C. R., Biagini, E., Blom, N. A., de Boer, R. A., De Winter, T., Elliott, P. M.,

Flather, M., Garcia-Pavia, P., Haugaa, K. H., Ingles, J., Jurcut, R. O., Klaassen, S., Limongelli, G., Loeys, B., Mogensen, J., Olivotto, I., Pantazis, A., Sharma, S., Van Tintelen, J. P., Ware, J. S., Kaski, J. P., & ESC Scientific Document Group. 2023 ESC Guidelines for the management of cardiomyopathies: Developed by the task force on the management of cardiomyopathies of the European Society of Cardiology (ESC). *European Heart Journal* **44,** 3503–3626 (2023).

5. McDonagh, T. A., Metra, M., Adamo, M., Gardner, R. S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak, M., Lam, C. S. P., Lyon, A. R., McMurray, J. J. V., Mebazaa, A., Mindham, R., Muneretto, C., Francesco Piepoli, M., Price, S., Rosano, G. M. C., Ruschitzka, F., Kathrine Skibelund, A., ESC Scientific Document Group, de Boer, R. A., Christian Schulze, P., Abdelhamid, M., Aboyans, V., Adamopoulos, S., Anker, S. D., Arbelo, E., Asteggiano, R., Bauersachs, J., Bayes-Genis, A., Borger, M. A., Budts, W., Cikes, M., Damman, K., Delgado, V., Dendale, P., Dilaveris, P., Drexel, H., Ezekowitz, J., Falk, V., Fauchier, L., Filippatos, G., Fraser, A., Frey, N., Gale, C. P., Gustafsson, F., Harris, J., Iung, B., Janssens, S., Jessup, M., Konradi, A., Kotecha, D., Lambrinou, E., Lancellotti, P., Landmesser, U., Leclercq, C., Lewis, B. S., Leyva, F., Linhart, A., Løchen, M.-L., Lund, L. H., Mancini, D., Masip, J., Milicic, D., Mueller, C., Nef, H., Nielsen, J.-C., Neubeck, L., Noutsias, M., Petersen, S. E., Sonia Petronio, A., Ponikowski, P., Prescott, E., Rakisheva, A., Richter, D. J., Schlyakhto, E., Seferovic, P., Senni, M., Sitges, M., Sousa-Uva, M., Tocchetti, C. G., Touyz, R. M., Tschoepe, C., Waltenberger, J., Adamo, M., Baumbach, A., Böhm, M., Burri, H., Čelutkienė, J., Chioncel, O., Cleland, J. G. F., Coats, A. J. S., Crespo-Leiro, M. G., Farmakis, D., Gardner, R. S., Gilard, M., Heymans, S., Hoes, A. W., Jaarsma, T., Jankowska, E. A., Lainscak,

M., Lam, C. S. P., Lyon, A. R., McMurray, J. J. V., Mebazaa, A., Mindham, R., Muneretto, C., Piepoli, M. F., Price, S., Rosano, G. M. C., Ruschitzka, F. & Skibelund, A. K. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *European Heart Journal* **42,** 3599–3726 (2021).

6. von Knobelsdorff-Brenkenhoff, F. & Schulz-Menger, J. Cardiovascular magnetic resonance in the guidelines of the European Society of Cardiology: a comprehensive summary and update. *Journal of Cardiovascular Magnetic Resonance* **25,** 42 (2023).
7. Gardner, B. I., Bingham, S. E., Allen, M. R., Blatter, D. D. & Anderson, J. L. Cardiac magnetic resonance versus transthoracic echocardiography for the assessment of cardiac volumes and regional function after myocardial infarction: an intrasubject comparison using simultaneous intrasubject recordings. *Cardiovascular Ultrasound* **7,** 38 (2009).
8. Stacey, R. B. & Hundley, W. G. The Role of Cardiovascular Magnetic Resonance (CMR) and Computed Tomography (CCT) in Facilitating Heart Failure Management. *Curr Treat Options Cardiovasc Med* **15,** 373–386 (2013).
9. Ansorge, R. & Graves, M. *The Physics and Mathematics of MRI*. (Morgan & Claypool Publishers, 2016). at <https://iopscience.iop.org/book/mono/978-1-6817-4068-3>
10. Lipton, M. L. *Totally Accessible MRI*. (Springer, 2008). doi:10.1007/978-0-387-48896-7
11. Weishaupt, D., Köchli, V. D. & Marincek, B. *Wie funktioniert MRI?: Eine Einführung in Physik und Funktionsweise der Magnetresonanzbildgebung*. (Springer, 2014). doi:10.1007/978-3-642-41616-3
12. Alfudhili, K., Masci, P. G., Delacoste, J., Ledoux, J.-B., Berchier, G., Dunet, V., Qanadli, S. D., Schwitter, J. & Beigelman-Aubry, C. Current artefacts in cardiac and chest magnetic resonance imaging: tips and tricks. *Br J Radiol* **89,** 20150987 (2016).

13. Larson, A. C., Kellman, P., Arai, A., Hirsch, G. A., McVeigh, E., Li, D. & Simonetti, O. P. Preliminary Investigation of Respiratory Self-Gating for Free-Breathing Segmented Cine MRI. *Magn Reson Med* **53,** 159–168 (2005).

14. Zhou, R., Huang, W., Yang, Y., Chen, X., Weller, D. S., Kramer, C. M., Kozerke, S. & Salerno, M. Simple motion correction strategy reduces respiratory-induced motion artifacts for k-t accelerated and compressed-sensing cardiovascular magnetic resonance perfusion imaging. *Journal of Cardiovascular Magnetic Resonance* **20,** 6 (2018).

15. Nacif, M. S., Zavodni, A., Kawel, N., Choi, E.-Y., Lima, J. A. C. & Bluemke, D. A. Cardiac magnetic resonance imaging and its electrocardiographs (ECG): tips and tricks. *Int J Cardiovasc Imaging* **28,** 1465–1475 (2012).

16. American Heart Association Writing Group on Myocardial Segmentation and Registration for Cardiac Imaging:, Cerqueira, M. D., Weissman, N. J., Dilsizian, V., Jacobs, A. K., Kaul, S., Laskey, W. K., Pennell, D. J., Rumberger, J. A., Ryan, T. & Verani, M. S. Standardized Myocardial Segmentation and Nomenclature for Tomographic Imaging of the Heart: A Statement for Healthcare Professionals From the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* **105,** 539–542 (2002).

17. Salerno, M., Sharif, B., Arheden, H., Kumar, A., Axel, L., Li, D. & Neubauer, S. Recent Advances in Cardiovascular Magnetic Resonance Techniques and Applications. *Circ Cardiovasc Imaging* **10,** e003951 (2017).

18. Schulz-Menger, J., Bluemke, D. A., Bremerich, J., Flamm, S. D., Fogel, M. A., Friedrich, M. G., Kim, R. J., von Knobelsdorff-Brenkenhoff, F., Kramer, C. M., Pennell, D. J., Plein, S. & Nagel, E. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update. *Journal of Cardiovascular Magnetic Resonance* **22,** 19 (2020).

19. Rigolli, M., Anandabaskaran, S., Christiansen, J. P. & Whalley, G. A. Bias associated with left ventricular quantification by multimodality imaging: a systematic review and meta-analysis. *Open Heart* **3,** e000388 (2016).

20. Messroghli, D. R., Moon, J. C., Ferreira, V. M., Grosse-Wortmann, L., He, T., Kellman, P., Mascherbauer, J., Nezafat, R., Salerno, M., Schelbert, E. B., Taylor, A. J., Thompson, R., Ugander, M., van Heeswijk, R. B. & Friedrich, M. G. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *Journal of Cardiovascular Magnetic Resonance* **19,** 75 (2017).

21. Salerno, M. & Kramer, C. M. Advances in Parametric Mapping With CMR Imaging. *JACC: Cardiovascular Imaging* **6,** 806–822 (2013).

22. Ogier, A. C., Bustin, A., Cochet, H., Schwitter, J. & van Heeswijk, R. B. The Road Toward Reproducibility of Parametric Mapping of the Heart: A Technical Review. *Frontiers in Cardiovascular Medicine* **9,** (2022).

23. Siemens Healthineers AG. MyoMaps. (2023). at <https://www.siemens-healthineers.com/magnetic-resonance-imaging/options-and-upgrades/clinical-applications/myomaps>

24. Koninklijke Philips N.V. CardiacQuant MR clinical application. *Philips* (2023). at <https://www.philips.co.uk/healthcare/product/HCNMRB808/cardiacquant-mr-clinical-application>

25. Karur, G. R. & Hanneman, K. Cardiac MRI T1, T2, and T2* Mapping in Clinical Practice. *Advances in Clinical Radiology* **1,** 27–41 (2019).

26. Warnica, W., Al-Arnawoot, A., Stanimirovic, A., Thavendiranathan, P., Wald, R. M., Pakkal, M., Karur, G. R., Wintersperger, B. J., Rac, V. & Hanneman, K. Clinical

Impact of Cardiac MRI T1 and T2 Parametric Mapping in Patients with Suspected Cardiomyopathy. *Radiology* **305,** 319–326 (2022).

27. Kellman, P., Wilson, J. R., Xue, H., Ugander, M. & Arai, A. E. Extracellular volume fraction mapping in the myocardium, part 1: evaluation of an automated method. *Journal of Cardiovascular Magnetic Resonance* **14,** 63 (2012).

28. Kellman, P., Wilson, J. R., Xue, H., Bandettini, W. P., Shanbhag, S. M., Druey, K. M., Ugander, M. & Arai, A. E. Extracellular volume fraction mapping in the myocardium, part 2: initial clinical experience. *Journal of Cardiovascular Magnetic Resonance* **14,** 64 (2012).

29. Messroghli, D. R., Radjenovic, A., Kozerke, S., Higgins, D. M., Sivananthan, M. U. & Ridgway, J. P. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. *Magnetic Resonance in Medicine* **52,** 141–146 (2004).

30. Piechnik, S. K., Ferreira, V. M., Dall'Armellina, E., Cochlin, L. E., Greiser, A., Neubauer, S. & Robson, M. D. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. *Journal of Cardiovascular Magnetic Resonance* **12,** 69 (2010).

31. Chow, K., Flewitt, J. A., Green, J. D., Pagano, J. J., Friedrich, M. G. & Thompson, R. B. Saturation recovery single-shot acquisition (SASHA) for myocardial T1 mapping. *Magnetic Resonance in Medicine* **71,** 2082–2095 (2014).

32. Heidenreich, J. F., Weng, A. M., Donhauser, J., Greiser, A., Chow, K., Nordbeck, P., Bley, T. A. & Köstler, H. T1- and ECV-mapping in clinical routine at 3 T: differences between MOLLI, ShMOLLI and SASHA. *BMC Med Imaging* **19,** (2019).

33. Kellman, P. & Hansen, M. S. T1-mapping in the heart: accuracy and precision. *Journal of Cardiovascular Magnetic Resonance* **16,** 2 (2014).

34. Taylor, A. J., Salerno, M., Dharmakumar, R. & Jerosch-Herold, M. T1 Mapping: Basic Techniques and Clinical Applications. *JACC: Cardiovascular Imaging* **9,** 67–81 (2016).

35. Pan, J., Hamdi, M., Huang, W., Hammernik, K., Kuestner, T. & Rueckert, D. Unrolled and rapid motion-compensated reconstruction for cardiac CINE MRI. *Medical Image Analysis* **91,** 103017 (2024).

36. Gonzales, R. A., Zhang, Q., Papież, B. W., Werys, K., Lukaschuk, E., Popescu, I. A., Burrage, M. K., Shanmuganathan, M., Ferreira, V. M. & Piechnik, S. K. MOCOnet: Robust Motion Correction of Cardiovascular Magnetic Resonance T1 Mapping Using Convolutional Neural Networks. *Front. Cardiovasc. Med.* **8,** (2021).

37. Xue, H., Shah, S., Greiser, A., Guetter, C., Littmann, A., Jolly, M.-P., Arai, A. E., Zuehlsdorff, S., Guehring, J. & Kellman, P. Motion correction for myocardial T1 mapping using image registration with synthetic image estimation. *Magn Reson Med* **67,** 1644–1655 (2012).

38. Leiner, T., Rueckert, D., Suinesiaputra, A., Baeßler, B., Nezafat, R., Išgum, I. & Young, A. A. Machine learning in cardiovascular magnetic resonance: basic concepts and applications. *Journal of Cardiovascular Magnetic Resonance* **21,** 61 (2019).

39. Roy, C., Slimani, A., de Meester, C., Amzulescu, M., Pasquet, A., Vancraeynest, D., Vanoverschelde, J.-L., Pouleur, A.-C. & Gerber, B. L. Age and sex corrected normal reference values of T1, T2 T2* and ECV in healthy subjects at 3T CMR. *Journal of Cardiovascular Magnetic Resonance* **19,** 72 (2017).

40. Dabir, D., Child, N., Kalra, A., Rogers, T., Gebker, R., Jabbour, A., Plein, S., Yu, C.-Y., Otton, J., Kidambi, A., McDiarmid, A., Broadbent, D., Higgins, D. M., Schnackenburg, B., Foote, L., Cummins, C., Nagel, E. & Puntmann, V. O. Reference values for

healthy human myocardium using a T1 mapping methodology: results from the International T1 Multicenter cardiovascular magnetic resonance study. *Journal of Cardiovascular Magnetic Resonance* **16,** 69 (2014).

41. Roujol, S., Weingärtner, S., Foppa, M., Chow, K., Kawaji, K., Ngo, L. H., Kellman, P., Manning, W. J., Thompson, R. B. & Nezafat, R. Accuracy, precision, and reproducibility of four T1 mapping sequences: a head-to-head comparison of MOLLI, ShMOLLI, SASHA, and SAPPHIRE. *Radiology* **272,** 683–689 (2014).

42. Child, N., Suna, G., Dabir, D., Yap, M.-L., Rogers, T., Kathirgamanathan, M., Arroyo-Ucar, E., Hinojar, R., Mahmoud, I., Young, C., Wendler, O., Mayr, M., Sandhu, B., Morton, G., Muhly-Reinholz, M., Dimmeler, S., Nagel, E. & Puntmann, V. O. Comparison of MOLLI, shMOLLLI, and SASHA in discrimination between health and disease and relationship with histologically derived collagen volume fraction. *European Heart Journal - Cardiovascular Imaging* **19,** 768–776 (2018).

43. Zange, L., Muehlberg, F., Blaszczyk, E., Schwenke, S., Traber, J., Funk, S. & Schulz-Menger, J. Quantification in cardiovascular magnetic resonance: agreement of software from three different vendors on assessment of left ventricular function, 2D flow and parametric mapping. *Journal of Cardiovascular Magnetic Resonance* **21,** 12 (2019).

44. McAlindon, E., Harris, J., Mathias, H., Marsden, D., Manghat, N., Reeves, B. & Bucciarelli-Ducci, C. CMR endpoints for clinical trials: impact of operator experience on the accuracy of image analysis. *Journal of Cardiovascular Magnetic Resonance* **14,** P38 (2012).

45. Carapella, V., Puchta, H., Lukaschuk, E., Marini, C., Werys, K., Neubauer, S., Ferreira, V. M. & Piechnik, S. K. Standardized image post-processing of cardiovascular

magnetic resonance T1-mapping reduces variability and improves accuracy and consistency in myocardial tissue characterization. *International Journal of Cardiology* **298,** 128–134 (2020).

46. Hann, E., Popescu, I. A., Zhang, Q., Gonzales, R. A., Barutçu, A., Neubauer, S., Ferreira, V. M. & Piechnik, S. K. Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Med Image Anal* **71,** 102029 (2021).

47. Puyol-Antón, E., Ruijsink, B., Baumgartner, C. F., Masci, P.-G., Sinclair, M., Konukoglu, E., Razavi, R. & King, A. P. Automated quantification of myocardial tissue characteristics from native T1 mapping using neural networks with uncertainty-based quality-control. *Journal of Cardiovascular Magnetic Resonance* **22,** 60 (2020).

48. Farrag, N. A., Lochbihler, A., White, J. A. & Ukwatta, E. Evaluation of fully automated myocardial segmentation techniques in native and contrast-enhanced T1-mapping cardiovascular magnetic resonance images using fully convolutional neural networks. *Med Phys* **48,** 215–226 (2021).

49. Merlo, M., Gagno, G., Baritussio, A., Bauce, B., Biagini, E., Canepa, M., Cipriani, A., Castelletti, S., Dellegrottaglie, S., Guaricci, A. I., Imazio, M., Limongelli, G., Musumeci, M. B., Parisi, V., Pica, S., Pontone, G., Todiere, G., Torlasco, C., Basso, C., Sinagra, G., Filardi, P. P., Indolfi, C., Autore, C. & Barison, A. Clinical application of CMR in cardiomyopathies: evolving concepts and techniques. *Heart Fail Rev* **28,** 77–95 (2023).

50. Gröschel, J., Trauzeddel, R.-F., Müller, M., von Knobelsdorff-Brenkenhoff, F., Viezzer, D., Hadler, T., Blaszczyk, E., Daud, E. & Schulz-Menger, J. Multi-site compar-

ison of parametric T1 and T2 mapping: healthy travelling volunteers in the Berlin research network for cardiovascular magnetic resonance (BER-CMR). *Journal of Cardiovascular Magnetic Resonance* **25,** 47 (2023).

51. Kranzusch, R., aus dem Siepen, F., Wiesemann, S., Zange, L., Jeuthe, S., Ferreira da Silva, T., Kuehne, T., Pieske, B., Tillmanns, C., Friedrich, M. G., Schulz-Menger, J. & Messroghli, D. R. Z-score mapping for standardized analysis and reporting of cardiovascular magnetic resonance modified Look-Locker inversion recovery (MOLLI) T1 data: Normal behavior and validation in patients with amyloidosis. *Journal of Cardiovascular Magnetic Resonance* **22,** 6 (2020).

52. Viezzer, D., Hadler, T., Gröschel, J., Ammann, C., Blaszczyk, E., Kolbitsch, C., Hufnagel, S., Kranzusch-Groß, R., Lange, S. & Schulz-Menger, J. Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept. *eBioMedicine* **102,** (2024).

53. Viezzer, D. S., Hadler, T. & Ammann, C. MARISSA - Magnetic Resonance Imaging Software for Standardization. (2024). doi:10.5281/zenodo.10640021

54. Viezzer, D., Hadler, T., Ammann, C., Blaszczyk, E., Fenski, M., Grandy, T. H., Wetzl, J., Lange, S. & Schulz-Menger, J. Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance. *Sci Rep* **13,** 2103 (2023).

55. The Medical Imaging Technology Association (MITA), a division of NEMA. NEMA PS3 / ISO 12052, Digital Imaging and Communications in Medicine (DICOM) Standard. at <http://www.dicomstandard.org/>

56. Hadler, T., Wetzl, J., Lange, S., Geppert, C., Fenski, M., Abazi, E., Gröschel, J., Ammann, C., Wenson, F., Töpper, A., Däuber, S. & Schulz-Menger, J. Introduction of

Lazy Luna an automatic software-driven multilevel comparison of ventricular function quantification in cardiovascular magnetic resonance imaging. *Sci Rep* **12,** 6629 (2022).

57. StatisticsHowTo.com. Statistics Definitions in Plain English with Examples. *Statistics How To* (2023). at <https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/>

58. J.Salkind, N. *Encyclopedia of Research Design*. (SAGE Publications, Inc., 2010). doi:10.4135/9781412961288

59. Turner, D. P. & Deng, H. Overview of Common Statistical Tests and their Assumptions. *Headache: The Journal of Head and Face Pain* **60,** 826–832 (2020).

60. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]* (2015). at <http://arxiv.org/abs/1505.04597>

61. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* (2017). at <http://arxiv.org/abs/1412.6980>

62. Jadon, S. A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)* 1–7 (2020). doi:10.1109/CIBCB48159.2020.9277638

63. Chollet, F. & others. Keras: Deep Learning for humans. (2015). at <https://keras.io/>

64. Smith, L. N. Cyclical Learning Rates for Training Neural Networks. Preprint at https://doi.org/10.48550/arXiv.1506.01186 (2017)

65. Prechelt, L. in *Neural Networks: Tricks of the Trade: Second Edition* (eds. Montavon, G., Orr, G. B. & Müller, K.-R.) 53–67 (Springer, 2012). doi:10.1007/978-3-642-35289-8_5

66. Muehlberg, F., Funk, S., Zange, L., von Knobelsdorff-Brenkenhoff, F., Blaszczyk, E., Schulz, A., Ghani, S., Reichardt, A., Reichardt, P. & Schulz-Menger, J. Native myocardial T1 time can predict development of subsequent anthracycline-induced cardiomyopathy. *ESC Heart Failure* **5,** 620–629 (2018).

67. Birukov, A., Wiesemann, S., Golic, M., Balogh, A., Marko, L., Rakova, N., Wilck, N., Blaszczyk, E., Lim, C., Weiss, S., Kräker, K., Haase, N., Frolova, A., Jørgensen, J. S., Daub, S., Müller, D. N., Herse, F., Schulz-Menger, J. & Dechend, R. Myocardial Evaluation of Post-Preeclamptic Women by CMR: Is Early Risk Stratification Possible? *JACC: Cardiovascular Imaging* **13,** 1291–1293 (2020).

68. Blaszczyk, E., Lim, C., Kellman, P., Schmacht, L., Gröschel, J., Spuler, S. & Schulz-Menger, J. Progressive myocardial injury in myotonic dystrophy type II and facioscapulohumeral muscular dystrophy 1: a cardiovascular magnetic resonance follow-up study. *Journal of Cardiovascular Magnetic Resonance* **23,** 130 (2021).

69. Stengl, H., Ganeshan, R., Hellwig, S., Blaszczyk, E., Fiebach, J. B., Nolte, C. H., Bauer, A., Schulz-Menger, J., Endres, M. & Scheitz, J. F. Cardiomyocyte Injury Following Acute Ischemic Stroke: Protocol for a Prospective Observational Cohort Study. *JMIR Res Protoc* **10,** e24186 (2021).

70. Shorten, C. & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* **6,** 60 (2019).

71. Martin Bland, J. & Altman, DouglasG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet* **327,** 307–310 (1986).

72. Taylor, R. Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography* **6,** 35–39 (1990).

73. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. & Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467 [cs]* (2016). at <http://arxiv.org/abs/1603.04467>

74. SciPy Developers. scipy: Fundamental algorithms for scientific computing in Python. (2024). at <https://scipy.org/>

75. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* **9,** 90–95 (2007).

76. GDAL/OGR contributors. GDAL: GDAL: Geospatial Data Abstraction Library. (2024). at <http://www.gdal.org>

77. geopy contributors. geopy: Python Geocoding Toolbox. (2018). at <https://github.com/geopy/geopy>

78. Collette, A. & contributors. h5py: Read and write HDF5 files from Python. (2008). at <https://www.h5py.org/>

79. Foord, M. & the mock team. mock: Rolling backport of unittest.mock for all Pythons. (2013). at <http://mock.readthedocs.org/en/latest/>

80. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C. & Oliphant, T. E. Array programming with NumPy. *Nature* **585,** 357–362 (2020).

81. Bradski, G. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2008).

82. openpyxl. openpyxl: A Python library to read/write Excel 2010 xlsx/xlsm files. (2010). at <https://openpyxl.readthedocs.io>

83. Reback, J., jbrockmendel, McKinney, W., Bossche, J. V. den, Augspurger, T., Cloud, P., Hawkins, S., gfyoung, Roeschke, M., Sinhrks, Klein, A., Petersen, T., Tratner, J., She, C., Ayd, W., Hoefler, P., Naveh, S., Garcia, M., Schendel, J., Hayden, A., Saxton, D., Darbyshire, J. H. M., Shadrach, R., Gorelli, M. E., Li, F., Zeitlin, M., Jancauskas, V., McMaster, A., Battiston, P. & Seabold, S. pandas-dev/pandas: Pandas 1.3.4. (2021). doi:10.5281/zenodo.5574486

84. Clark, J. A. & contributors. pillow: Python Imaging Library (Fork). (2024). at <https://python-pillow.org>

85. The pip developers. pip: The PyPA recommended tool for installing Python packages. (2020). at <https://pip.pypa.io/>

86. Jansen, F. & Custódio, B. M. polyline: A Python implementation of Google's Encoded Polyline Algorithm Format. (2023). at <https://github.com/frederickjansen/polyline>

87. Mason, D. & pydicom contributors. pydicom: A pure Python package for reading and writing DICOM data. (2020). at <https://github.com/pydicom/pydicom>

88. Gillies, S. & contributors. rasterio: Fast and direct raster I/O for use with Numpy and SciPy. (2021). at <https://github.com/rasterio/rasterio>

89. scikit-fuzzy team. scikit-fuzzy: Fuzzy logic toolkit for SciPy. (2012). at <https://pypi.python.org/pypi/scikit-fuzzy>

90. Walt, S. van der, Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E. & Yu, T. scikit-image: image processing in Python. *PeerJ* **2,** e453 (2014).

91. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. & Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12,** 2825–2830 (2011).

92. Gillies, S., van der Wel, C. & Shapely contributors. shapely: Manipulation and analysis of geometric objects. (2022). at <https://github.com/shapely/shapely>

93. Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python. in 92–96 (2010). doi:10.25080/Majora-92bf1922-011

94. McNamara, J. XlsxWriter: A Python module for creating Excel XLSX files. (2024). at <https://github.com/jmcnamara/XlsxWriter>

95. Zoomer Analytics GmbH. xlwings: Make Excel fly: Interact with Excel from Python and vice versa. (2024). at <https://www.xlwings.org>

96. Zou, K. H., O'Malley, A. J. & Mauri, L. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation* **115,** 654–657 (2007).

97. Power, M., Fell, G. & Wright, M. Principles for high-quality, high-value testing. *BMJ Evidence-Based Medicine* **18,** 5–10 (2013).

98. Hipp, R. SQLite. (2024). at <https://www.sqlite.org/index.html>

99. Demir, A., Wiesemann, S., Erley, J., Schmitter, S., Trauzeddel, R. F., Pieske, B., Hansmann, J., Kelle, S. & Schulz-Menger, J. Traveling Volunteers: A Multi-Vendor, Multi-Center Study on Reproducibility and Comparability of 4D Flow Derived Aortic Hemodynamics in Cardiovascular Magnetic Resonance. *J Magn Reson Imaging* **55,** 211–222 (2022).

100. Schmacht, L., Traber, J., Grieben, U., Utz, W., Dieringer, M. A., Kellman, P., Blaszczyk, E., von Knobelsdorff-Brenkenhoff, F., Spuler, S. & Schulz-Menger, J. Cardiac Involvement in Myotonic Dystrophy Type 2 Patients With Preserved Ejection Fraction. *Circulation: Cardiovascular Imaging* **9,** e004615 (2016).

101. von Knobelsdorff-Brenkenhoff, F., Schüler, J., Dogangüzel, S., Dieringer, M. A., Rudolph, A., Greiser, A., Kellman, P. & Schulz-Menger, J. Detection and Monitoring of Acute Myocarditis Applying Quantitative Cardiovascular Magnetic Resonance. *Circ Cardiovasc Imaging* **10,** e005242 (2017).

102. Blaszczyk, E., Töpper, A., Schmacht, L., Wanke, F., Greiser, A., Schulz-Menger, J. & von Knobelsdorff-Brenkenhoff, F. Influence of spatial resolution and contrast agent dosage on myocardial T1 relaxation times. *MAGMA* **30,** 85–91 (2017).

103. Liang, L., Wang, X., Yu, Y., Zhang, Y., Liu, J., Chen, M., Zhang, L. & Jiang, T. T1 Mapping and Extracellular Volume in Cardiomyopathy Showing Left Ventricular Hypertrophy: Differentiation Between Hypertrophic Cardiomyopathy and Hypertensive Heart Disease. *Int J Gen Med* **15,** 4163–4173 (2022).

104. Qin, L., Min, J., Chen, C., Zhu, L., Gu, S., Zhou, M., Yang, W. & Yan, F. Incremental Values of T1 Mapping in the Prediction of Sudden Cardiac Death Risk in Hypertrophic Cardiomyopathy: A Comparison With Two Guidelines. *Frontiers in Cardiovascular Medicine* **8,** (2021).

105. Lavall, D., Vosshage, N. H., Geßner, R., Stöbe, S., Ebel, S., Denecke, T., Hagendorff, A. & Laufs, U. Native T1 mapping for the diagnosis of cardiac amyloidosis in patients with left ventricular hypertrophy. *Clin Res Cardiol* **112,** 334–342 (2023).

106. Neubauer, S., Kolm, P., Ho, C. Y., Kwong, R. Y., Desai, M. Y., Dolman, S. F., Appelbaum, E., Desvigne-Nickens, P., DiMarco, J. P., Friedrich, M. G., Geller, N., Harper, A. R., Jarolim, P., Jerosch-Herold, M., Kim, D.-Y., Maron, M. S., Schulz-Menger,

J., Piechnik, S. K., Thomson, K., Zhang, C., Watkins, H., Weintraub, W. S., Kramer, C. M., Mahmod, M., Jacoby, D., White, J., Chiribiri, A., Helms, A., Choudhury, L., Michels, M., Bradlow, W., Salerno, M., Heitner, S., Prasad, S., Mohiddin, S., Swoboda, P., Mahrholdt, H., Bucciarelli-Ducci, C., Weinsaft, J., Kim, H., McCann, G., van Rossum, A., Williamson, E., Flett, A., Dawson, D., Mongeon, F. P., Olivotto, I., Crean, A., Owens, A., Anderson, L., Biagini, E., Newby, D., Berry, C., Kim, B., Larose, E., Abraham, T., Sherrid, M., Nagueh, S., Rimoldi, O., Elstein, E. & Autore, C. Distinct Subgroups in Hypertrophic Cardiomyopathy in the NHLBI HCM Registry. *Journal of the American College of Cardiology* **74,** 2333–2345 (2019).

107. Harper, A. R., Goel, A., Grace, C., Thomson, K. L., Petersen, S. E., Xu, X., Waring, A., Ormondroyd, E., Kramer, C. M., Ho, C. Y., Neubauer, S., Tadros, R., Ware, J. S., Bezzina, C. R., Farrall, M. & Watkins, H. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat Genet* **53,** 135–142 (2021).

108. Smith, G. in *Essential Statistics, Regression, and Econometrics (Second Edition)* (ed. Smith, G.) 219–259 (Academic Press, 2015). doi:10.1016/B978-0-12-803459-0.00008-X

109. Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. & Lin, C.-J. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research* **9,** 1871–1874 (2008).

110. Breiman, L. Random Forests. *Machine Learning* **45,** 5–32 (2001).

111. Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach Learn* **63,** 3–42 (2006).

112. Riverbank Computing Limited. PyQt5: Python bindings for the Qt cross platform application toolkit. (2024). at <https://www.riverbankcomputing.com/software/pyqt/>

113. The Qt Company Ltd. Qt Designer Manual. (2023). at <https://doc.qt.io/qt-5/qtdesigner-manual.html>

114. Balciunas, P. fnv: FNV hash functions. (2016). at <https://github.com/povilasb/pyfnv>

115. Gautier, L. rpy2: Python interface to the R language (embedded R). (2024). at <https://rpy2.github.io>

116. Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software* **6,** 3021 (2021).

117. Bayer, M. in *The Architecture of Open Source Applications Volume II: Structure, Scale, and a Few More Fearless Hacks* at <http://aosabook.org/en/sqlalchemy.html>

118. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L. & Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. Preprint at https://doi.org/10.48550/arXiv.2102.04306 (2021)

119. Peng, P., Lekadir, K., Gooya, A., Shao, L., Petersen, S. E. & Frangi, A. F. A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *MAGMA* **29,** 155–195 (2016).

120. Luca, A. R., Ursuleanu, T. F., Gheorghe, L., Grigorovici, R., Iancu, S., Hlusneac, M. & Grigorovici, A. Impact of quality, type and volume of data used by deep learning models in the analysis of medical images. *Informatics in Medicine Unlocked* **29,** 100911 (2022).

121. Dodge, S. & Karam, L. *Understanding How Image Quality Affects Deep Neural Networks*. (arXiv, 2016). at <http://arxiv.org/abs/1604.04004>

122. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., Lee, A. M., Aung, N., Lukaschuk, E., Sanghvi, M. M., Zemrak, F., Fung, K., Paiva, J. M., Carapella, V., Kim, Y. J., Suzuki, H., Kainz, B., Matthews, P. M., Petersen, S. E., Piechnik, S. K.,

Neubauer, S., Glocker, B. & Rueckert, D. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance* **20,** 65 (2018).

123. Ammann, C., Hadler, T., Gröschel, J., Kolbitsch, C. & Schulz-Menger, J. Multilevel comparison of deep learning models for function quantification in cardiovascular magnetic resonance: On the redundancy of architectural variations. *Front Cardiovasc Med* **10,** 1118499 (2023).

124. Kaur, J. & Singh, W. A systematic review of object detection from images using deep learning. *Multimed Tools Appl* **83,** 12253–12338 (2024).

125. Zhao, Z.-Q., Zheng, P., Xu, S. & Wu, X. Object Detection with Deep Learning: A Review. *arXiv:1807.05511 [cs]* (2019). at <http://arxiv.org/abs/1807.05511>

126. Niu, Y., Qin, L. & Wang, X. Myocardium Detection by Deep SSAE Feature and Within-Class Neighborhood Preserved Support Vector Classifier and Regressor. *Sensors (Basel)* **19,** 1766 (2019).

127. Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P. & Johansen, H. D. *DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation*. (arXiv, 2020). doi:10.48550/arXiv.2006.04868

128. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18,** 203–211 (2021).

129. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. in *Advances in Neural Information Processing Systems* **25,** (Curran Associates, Inc., 2012).

130. Bergstra, J. & Bengio, Y. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research* **13,** 281–305 (2012).

131. Fadil, H., Totman, J. J., Hausenloy, D. J., Ho, H.-H., Joseph, P., Low, A. F.-H., Richards, A. M., Chan, M. Y. & Marchesseau, S. A deep learning pipeline for automatic analysis of multi-scan cardiovascular magnetic resonance. *J Cardiovasc Magn Reson* **23,** 47 (2021).

132. Shaw, M., Ojha, V., Ganga, K. P., Malhi, A. S., Chandrashekhara, S. H., Kumar, S., Khan, M. A., Jagia, P. & Sharma, S. Reference values of myocardial native T1 and T2 mapping values in normal Indian population at 1.5 Tesla scanner. *Int J Cardiovasc Imaging* **38,** 2403–2411 (2022).

133. Abdullah, H., Abdul Kadir, A., Sha'rani, M. A., Tan, S. H. & Othman, N. Pilot study to establish the myocardium native T1 mapping reference values in 1.5 Tesla cardiac magnetic resonance of healthy Malaysians. *Egyptian Journal of Radiology and Nuclear Medicine* **54,** 98 (2023).

134. Böttcher, B., Lorbeer, R., Stöcklein, S., Beller, E., Lang, C. I., Weber, M.-A. & Meinel, F. G. Global and Regional Test–Retest Reproducibility of Native T1 and T2 Mapping in Cardiac Magnetic Resonance Imaging. *Journal of Magnetic Resonance Imaging* **54,** 1763–1772 (2021).

135. Granitz, M., Motloch, L. J., Granitz, C., Meissnitzer, M., Hitzl, W., Hergan, K. & Schlattau, A. Comparison of native myocardial T1 and T2 mapping at 1.5T and 3T in healthy volunteers : Reference values and clinical implications. *Wien Klin Wochenschr* **131,** 143–155 (2019).

136. Zhao, H., Huang, R., Jiang, M., Wang, W., Chai, Y., Liu, Q., Zhang, W., Han, Y., Yan, F., Lu, Q., Tao, Z., Wu, Q., Yue, J., Ma, J. & Pu, J. Myocardial Tissue-Level Characteristics of Adults With Metabolically Healthy Obesity. *JACC: Cardiovascular Imaging* **16,** 889–901 (2023).

137. Hufnagel, S., Metzner, S., Kerkering, K. M., Aigner, C. S., Kofler, A., Schulz-Menger, J., Schaeffter, T. & Kolbitsch, C. 3D model-based super-resolution motion-corrected cardiac T1 mapping. *Phys. Med. Biol.* **67,** 245008 (2022).

138. WHO Expert Committee on Physical Status : the Use and Interpretation of Anthropometry (1993 : Geneva, S. & Organization, W. H. *Physical status : the use of and interpretation of anthropometry , report of a WHO expert committee*. (World Health Organization, 1995). at <https://apps.who.int/iris/handle/10665/37003>

139. Soler, R., Méndez, C., Rodríguez, E., Barriales, R., Ochoa, J. P. & Monserrat, L. Phenotypes of hypertrophic cardiomyopathy. An illustrative review of MRI findings. *Insights Imaging* **9,** 1007–1020 (2018).

140. Baggiano, A., Boldrini, M., Martinez-Naharro, A., Kotecha, T., Petrie, A., Rezk, T., Gritti, M., Quarta, C., Knight, D. S., Wechalekar, A. D., Lachmann, H. J., Perlini, S., Pontone, G., Moon, J. C., Kellman, P., Gillmore, J. D., Hawkins, P. N. & Fontana, M. Noncontrast Magnetic Resonance for the Diagnosis of Cardiac Amyloidosis. *JACC: Cardiovascular Imaging* **13,** 69–80 (2020).

141. Nam, B. D., Kim, S. M., Jung, H. N., Kim, Y. & Choe, Y. H. Comparison of quantitative imaging parameters using cardiovascular magnetic resonance between cardiac amyloidosis and hypertrophic cardiomyopathy: inversion time scout versus T1 mapping. *Int J Cardiovasc Imaging* **34,** 1769–1777 (2018).

142. Martinez-Naharro, A., Kotecha, T., Norrington, K., Boldrini, M., Rezk, T., Quarta, C., Treibel, T. A., Whelan, C. J., Knight, D. S., Kellman, P., Ruberg, F. L., Gillmore, J. D., Moon, J. C., Hawkins, P. N. & Fontana, M. Native T1 and Extracellular Volume in Transthyretin Amyloidosis. *JACC: Cardiovascular Imaging* **12,** 810–819 (2019).

143. Popescu, I. A., Werys, K., Zhang, Q., Puchta, H., Hann, E., Lukaschuk, E., Ferreira, V. M. & Piechnik, S. K. Standardization of T1-mapping in cardiovascular magnetic

resonance using clustered structuring for benchmarking normal ranges. *International Journal of Cardiology* **326,** 220–225 (2021).

144. Kalapos, A., Szabó, L., Dohy, Z., Kiss, M., Merkely, B., Gyires-Tóth, B. & Vágó, H. Automated T1 and T2 mapping segmentation on cardiovascular magnetic resonance imaging using deep learning. *Front Cardiovasc Med* **10,** 1147581 (2023).

Statutory Declaration

“I, Darian Steven Viezzer, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic *Harmonization of confounding parameters to minimize induced biases on parametric T1 maps in cardiovascular magnetic resonance imaging* [*Harmonisierung von Einflussfaktoren zur Minimierung von induzierten Verzerrungseffekten bei parametrischen T1 Karten in der kardiovaskulären Magnetresonanztomographie*], independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me.”

_____ _____

Date Signature

Declaration of your own contribution to the publications

Darian Steven Viezzer contributed the following to the below listed publications according to the Contributor Roles Taxonomy (CRediT found at https://credit.niso.org):

Publication 1: **Viezzer, D.**, Hadler, T., Ammann, C., Blaszczyk, E., Fenski, M., Grandy, T. H., Wetzl, J., Lange, S. & Schulz-Menger, J., Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance, Scientific Reports, 2023

Contribution:
- Conceptualization
- Data Curation
- Formal Analysis
- Investigation
- Methodology
- Project Administration
- Software
- Validation
- Visualization (all tables, all figures & supplemental material drafted and edited)
- Writing (original draft & review & editing)

Publication 2: **Viezzer, D.**, Hadler, T., Gröschel, J., Ammann, C., Blaszczyk, E., Kolbitsch, C., Hufnagel, S., Kranzusch-Groß, R., Lange, S. & Schulz-Menger, J., Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept, eBioMedicine, 2024

Contribution:
- Conceptualization
- Data Curation
- Formal Analysis
- Investigation
- Methodology
- Project Administration
- Software
- Validation
- Visualization (all tables, all figures & supplemental material drafted and edited)
- Writing (original draft & review & editing)

_____

Signature, date and stamp of first supervising university professor / lecturer

_____

Signature of doctoral candidate

Printing copy(s) of the publication(s)

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance

Darian Viezzer[1,2,3], Thomas Hadler[1,2,3], Clemens Ammann[1,2,3], Edyta Blaszczyk[1,2,3], Maximilian Fenski[1,2,4], Thomas Hiroshi Grandy[1,2,4], Jens Wetzl[5], Steffen Lange[6] & Jeanette Schulz-Menger[1,2,3,4]✉

The manual and often time-consuming segmentation of the myocardium in cardiovascular magnetic resonance is increasingly automated using convolutional neural networks (CNNs). This study proposes a cascaded segmentation (CASEG) approach to improve automatic image segmentation quality. First, an object detection algorithm predicts a bounding box (BB) for the left ventricular myocardium whose 1.5 times enlargement defines the region of interest (ROI). Then, the ROI image section is fed into a U-Net based segmentation. Two CASEG variants were evaluated: one using the ROI cropped image solely (cropU) and the other using a 2-channel-image additionally containing the original BB image section (crinU). Both were compared to a classical U-Net segmentation (refU). All networks share the same hyperparameters and were tested on basal and midventricular slices of native and contrast enhanced (CE) MOLLI T1 maps. Dice Similarity Coefficient improved significantly ($p < 0.05$) in cropU and crinU compared to refU (81.06%, 81.22%, 72.79% for native and 80.70%, 79.18%, 71.41% for CE data), while no significant improvement ($p < 0.05$) was achieved in the mean absolute error of the T1 time (11.94 ms, 12.45 ms, 14.22 ms for native and 5.32 ms, 6.07 ms, 5.89 ms for CE data). In conclusion, CASEG provides an improved geometric concordance but needs further improvement in the quantitative outcome.

Cardiovascular magnetic resonance (CMR) is one of the most important non-invasive imaging modalities for risk stratification in cardiovascular diseases[1,2]. It enables the characterization of focal and diffuse changes in the myocardial tissue by quantitative techniques such as parametric mapping[3], which is considered as one of the most meaningful innovations in recent CMR developments[4,5]. While T2 mapping is used for the detection of myocardial edemas, T1 mapping is applied across multiple pathophysiological mechanisms and tissue characteristics[4]. A T1 map can be acquired before, referred as native, or after application of a contrast agent, referred to as contrast enhanced (CE)[6]. The latter is usually integrated in the calculation of an extracellular volume (ECV) map.

For diagnostics, the post-processing involves the segmentation of the myocardium as tissue of interest in order to obtain quantitative values[7]. This is usually performed in a manual and often time consuming manner[1,7–10].

[1]ECRC Experimental and Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Lindenberger Weg 80, 13125 Berlin, Germany. [2]Working Group on Cardiovascular Magnetic Resonance, Experimental and Clinical Research Center, a joint cooperation between the Charité – Universitätsmedizin Berlin and the Max-Delbrück-Center for Molecular Medicine, Berlin, Germany. [3]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany. [4]Department of Cardiology and Nephrology, Helios Hospital Berlin-Buch, Berlin, Germany. [5]Siemens Healthcare GmbH, Erlangen, Germany. [6]Faculty for Computer Sciences, Hochschule Darmstadt (University of Applied Sciences), Darmstadt, Germany. ✉email: jeanette.schulz-menger@charite.de

Furthermore, even experienced readers show intra-observer variability that results in considerable quantification uncertainty[11]. Current developments attempt to overcome both, the time consumption and the reproducibility uncertainty by using convolutional neural networks (CNNs) to automatically segment the tissue of interest. One of the most prominent CNN models in medical segmentation tasks is the U-Net[1] introduced by Ronneberger et al.[12]. While a standard U-Net already shows an average geometrical overlap above 70% with an expert segmentation[8–10,13], which is conventionally assumed as a good result[8], current development focuses on technical improvements for even better segmentation quality. At present, many strategies intend to advance the CNN models towards more complex framework structures[8,9] or integrating alternative architecture structures[14].

Parametric maps are images with pixel values representing the amplitude of a physical quantity. However, the vast majority of pixels contain superfluous background information. Object detection algorithms (ODAs) are used to find object regions of interest (ROI) in images by localizing a bounding box (BB) around those objects with the help of CNNs[15,16]. Consequently, ODAs can help to focus on a ROI in parametric mapping and thus, to reduce the amount of background information that is fed into an automatic segmentation network.

Although the detection of left ventricular myocardium by ODAs already exist for CINE images in CMR[16], its application on parametric T1 mapping and combination with automatically segmenting CNN procedures remains to the best of our knowledge unaddressed. Consequently, the aim of this study is to analyze the impact of input data enhancement on the segmentation quality in parametric T1 mapping by introducing an ODA as a preliminary processing step before the actual segmentation task. This coarse to fine segmentation procedure is named in the following as cascaded segmentation (CASEG).

Materials and methods

Dataset. A heterogeneous dataset of parametric T1 maps with corresponding manual reference segmentation from published[17–19] and on-going studies[20] was used. The inline T1 maps were either generated on a 1.5 T AvantoFit, a 3 T SkyraFit or a 3 T PrismaFit clinical magnetic resonance imaging scanner (all Siemens Healthcare, Erlangen, Germany) and were based on the MOLLI sequence using a 5(3)3 scheme for native and a 4(1)3(1)2 scheme for CE acquisitions.

Data from N = 403 participants (97 healthy volunteers and 306 patients) were used resulting in a total of M = 1438 parametric T1 maps, of which 1080 were native and 358 CE T1 maps. The difference in the quantity of native compared to CE T1 maps is due to the absence of CE measurements in some of the original studies. The dataset was randomly split per study set into 75% training, 10% validation and 15% test data. Table 1 shows an overview of the dataset and the amount for training, validation and testing. While some source studies in the dataset contained a full short axis T1 map stack, others only had three (basal, midventricular and apical), two (mostly basal and midventricular) or solely one (mostly midventricular) slice. The training and validation were done on all assigned T1 maps to assure for an advanced generalization of the segmentation network, whereas, the test dataset was restricted to midventricular and basal slices only as recommended by the society for cardiovascular magnetic resonance[4]. The reference segmentation was performed manually by experienced readers using the software cvi42 (Circle Cardiovascular Imaging, Calgary, Canada). The data acquisition and manual segmentation processing were performed in accordance with relevant guidelines and regulations. This study was approved by the local ethics committee of the Charité Universitätsmedizin Berlin (study ID: EA 1 253 21).

ODA. A CNN based ODA was used to detect a BB[15] that tightly fits the left ventricular myocardium in the parametric T1 map. During training, the ODA CNN behaved like a conventional segmentation CNN by providing the target BB data as binary mask. The predicted raw output of the ODA CNN model, in turn, needed a post-processing in order to represent a binary mask of a BB. First, the output was thresholded at a value of 0.5, then the largest connected component (LCC) was identified as the BB ROI. The minimum and maximum indeces of the LCC along both image axis defined the BB edges. These edges were finally converted into a BB binary mask.

A magnification factor to enlarge the BB was evaluated in order to securely cover the whole left ventricle within the ROI. The maximum occurring factor across the test data is assumed as suitable to guarantee for this. This factor was used to compensate uncertainties from the CNN based prediction while keeping the ROI small compared to the original image size.

CASEG. The basic idea of CASEG was the subsequent arrangement of two independent CNN models. The ODA served as a first coarse prediction of a potential ROI while the second CNN was applied on the focused image section in the ROI for the actual segmentation task and returned the final segmentation mask.

	Training	Validation	Testing	Total
Subjects (N)	313	35	55 (55)	403
Native T1 maps	849	91	140 (106)	1080
Contrast enhanced T1 maps	286	27	45 (33)	358
Total T1 maps (M)	1135	118	185 (139)	1438

Table 1. Overview of the complete dataset, the (numbers) in brackets denotes the number for midventricular and basal slices only that are used as test dataset in this study.

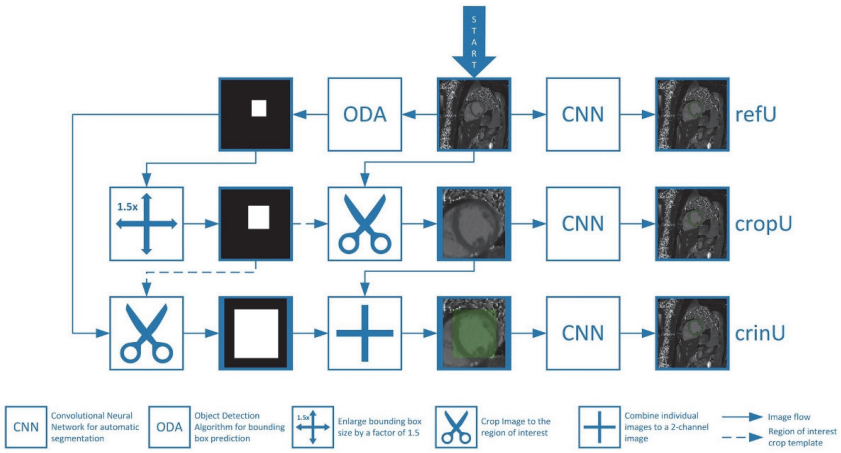

Figure 1. Processing pipelines for refU, cropU and crinU. Convolutional neural networks (CNNs) are used for the segmentation of the myocardium as tissue of interest. While refU directly uses the input image, cropU and crinU use the region of interest image section that belongs to the 1.5 times enlarged bounding box from an object detection algorithm (ODA). In contrast to cropU, crinU uses a two channel image with the second channel having the original predicted bounding box mask.

Due to the BB enlargement for the ROI definition, two potential CASEG pipelines (cropU and crinU) were evaluated as visualized in Fig. 1. In cropU, the input image for the secondary segmentation CNN was the original image cropped to the ROI image section. In contrast to that, the crinU worked in a similar manner and also considered the cropped image section. However, in crinU, the input image for the secondary segmentation CNN was extended by a second channel that comprises the unenlarged BB (shown as green overlay in Fig. 1). Considering cropU, an alternative implementation exists as visualized in Supplemental Material S1 by having an ODA that directly predicts an enlarged BB without the magnification factor processing step between ODA and segmentation network (cropU_A).

All three CASEG pipelines were compared to a reference pipeline (refU) that was equivalent to the secondary segmentation model without any preliminary ODA such that the original image was directly fed into the CNN.

Quality assurance. All CNN models including the ODA were implemented as U-Nets. As this study aimed to analyze the impact of input data enhancement by using a preliminary ODA, the same hyperparameters were used for all CNNs except for the number of input channels. The selected hyperparameters showed the best results in the reference refU. The models had 27 layers with 6 skip connections and the input size was $256 \times 256 \times 2$ for the secondary segmentation model in crinU and $256 \times 256 \times 1$ for all other U-Net models. A detailed U-Net structure overview is shown in the Supplemental Material S2. Consequently, the input images were resized and interpolated to the model specific size. The log-cosh-dice loss function[21] with an Adam optimizer[22] having a clip-norm of 0.001 was used. The batch size was set to 10 and the number of epochs was 1000 but an early stopping scheme[23] that ended the training after 50 epochs of no improvement with respect to the dice similarity coefficient metric in the validation data was used. The learning rate, which reflected the maximum learning rate value the Adam optimizer could capture, of initially 0.001 was halved every 25 epochs of no improvement.

During the training of the individual models, the training dataset was randomly augmented[24] with brightness adjustments, contrast adjustments, blurring, Gaussian random noise, salt and pepper noise, rotation, mirroring, axis downsampling and, for the refU and ODA only, a cropping of the image. Additionally, during training of the secondary segmentation CNN models in cropU and crinU, the detected BB was randomly shifted and resized by up to 5 pixels, whereas in 5% the optimal BB was used instead of the predicted one and in another 5% a failed BB detection was assumed in order to reflect potential detection uncertainties.

In case of a BB prediction failure, the ODA returned a binary image with only zero values reflecting no found BB. In such a case, cropU behaved similar to refU and used the original image as input. Further, crinU used the original image as first channel as well and kept the second channel zero valued. If the BB enlargement exceeded the image boundaries, the ROI was cropped at that boundary such that it never exceeded the original image.

For normalization of the input images, each input image channel was scaled to floating point values between zero and one. Further, while the input images were internally resized to the model specific size, the output was back transformed to the original input image size. To losslessly apply this and other geometric transformations, the segmentation masks were converted from binary pixel masks to vectorized contour objects allowing for geometrically precise transformation. The transformed structures were then rasterized back into pixel masks.

The primary domain of the output quality measure is of geometric nature. The geometric domain reflects the spatial similarity of two individual segmentations. For that reason, the Dice Similarity Coefficient (DSC) and the Hausdorff Distance (HD) were used as geometrical quality metrics. Both were supported visually with boxplots. As the DSC and HD were not normally distributed, non-parametric Friedman (across all models) and Wilcoxon (across refU and either CASEG pipeline) tests were used with a significance level of $p < 0.05$. Significance was assumed if both statistical tests were significant. Additionally, the enlarged BB detection was tested for the increment in the ratio of relevant pixels reflecting the foreground information to the total number of pixels in the image section compared to the original image. The ratio increment was tested with a Wilcoxon test for significance with a significance level of $p < 0.05$.

As parametric T1 maps provide clinically interpretable quantitative measurements, the quantitative domain was tested for the effect of the four segmentation approaches on the estimated average T1 time. The mean error (ME), mean absolute error (MAE) and root mean squared error (RMSE) were used in combination with the confidence interval (CI) to evaluate the quality in the quantitative domain. The CI were tested for remaining within the published equivalence margin derived from an intra-observer variability of native T1 maps, which is defined as the clinically acceptable deviation[11]. As the four models were evaluated on the same test dataset, the CIs were Bonferoni corrected[11]. The coefficient of variation (CV) with respect to the quantitative T1 error were additionally provided in order to analyze if the CE segmentations would remain in an adequate equivalence margin for CE data.

As the ME and MAE are not normally distributed, non-parametric Friedman and Wilcoxon tests were used for the comparison of these metrics with a significance level of $p < 0.05$. A lower RMSE was assumed to indicate an improvement.

Additionally, a correlation plot including linear regression, the Pearson Correlation (testing for linearity) and Kendall's Tau (testing for rank-order stability) were provided. While conventionally correlation coefficient values are assumed as weak if smaller than 0.35, moderate if up to 0.67, strong if up to 0.90 and very strong if above, the coefficient of determination (CoD, squared Pearson correlation coefficient), represents the amount of shared variance between two measures and thus may support the interpretation of the findings[25]. The correlation plot was complemented with a Bland–Altman plot in order to visualize the limits of agreement[26].

As T1 times of native and CE maps are on different scales, all analysis were separately performed on native and CE data.

Source code. All programming tasks were implemented in Python (version 3.8, Python Software Foundation, Beaverton, USA). The necessary Python libraries with their specific used version are listed in the Supplemental Material S3. The U-Nets were implemented using the Tensorflow[27] library (version 2.7.0) while statistics were calculated with the scipy[28] package (version 1.4.1) and plots were created with the matplotlib[29] library (version 3.5.2). The software includes a README file with a description for using the software. The user does not need to take care about processing between the ODA and the segmentation network as this is done automatically in the provided scripts.

Ethical approval. This study was approved by the local ethics committee of the Charité Universitätsmedizin Berlin as retrospective study (study ID: EA 1 253 21). The requirement for written informed consent was acquired during the original clinical studies and was therefore waived in this study due to its retrospective design as approved by the local ethics committee of the Charité Universitätsmedizin Berlin (study ID: EA 1 253 21).

Results

Numbers in the results are shown as mean ± standard deviation with their specific unit of measure.

ODA. The U-Net based ODA identified a BB of the left ventricle in all cases of the test dataset. Figure 2 shows respectively the best and worst cases for the BB prediction in regard of DSC and HD across native and CE test data. Throughout the whole test dataset, the predicted BBs resulted in a DSC of 93.09 ± 2.13% and 91.99 ± 2.80% and a HD of 3.95 ± 1.07 mm and 4.42 ± 2.40 mm for native and CE T1 maps respectively.

In order to securely cover the whole left ventricle with the BB across the test dataset, a magnification factor of at least 1.44 was necessary. Hence, cropU and crinU were set up with a rounded-up magnification factor of 1.50. In the Supplemental Material S4 the impact of the magnification factor on the average DSC results in the test dataset for cropU and crinU is shown. For magnification factors between 1.3 and 2.5, the results reached a performance plateau with minor fluctuations due to model training uncertainties.

Comparing the ratio of relevant foreground pixels to the total number of pixels, the ratio increased significantly ($p < 0.05$) to10.38 ± 3.27% and 10.76 ± 4.22% in the test dataset for native and CE maps as compared to 0.71 ± 0.35% and 0.62 ± 0.21% when using the ROI image section instead of the original image. Assuming an unenlarged perfectly fitting BB, the maximum reachable ratio would be 20.12 ± 6.67% and 19.78 ± 6.00% for native and CE test data respectively. In the training and validation dataset, the increment was similar. Detailed boxplots of the ratio of relevant pixels are provided in the Supplemental Material S5 separately for native and CE data in the training, validation and test datasets.

CASEG. Exemplary segmentation results for refU, cropU, crinU and cropU_A are shown in Fig. 3 with a good case across all four pipelines, a case improving in CASEG compared to refU and a rather poor case across all four pipelines respectively for native and CE T1 maps. In Table 2 detailed results of geometric and quantitative metrics are provided.

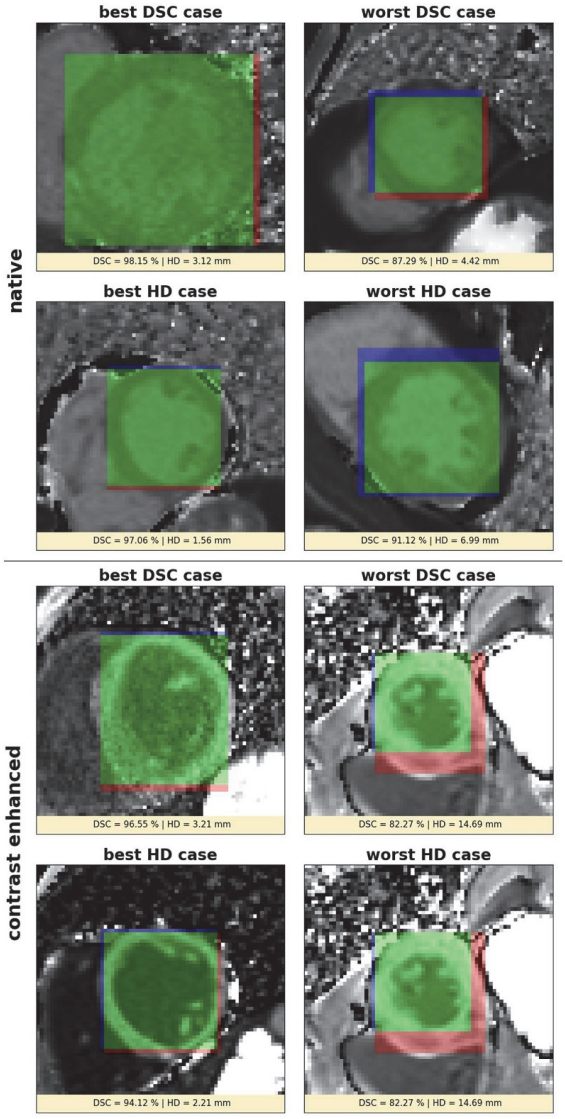

Figure 2. Example results of the object detection algorithm showing bounding boxes for the left ventricular myocardium. The upper block corresponds to native and the lower block to contrast enhanced data; respectively in each block the first row corresponds with respect to the Dice Similarity Coefficient (DSC) and the second row corresponds with respect to the Hausdorff Distance (HD) while the first column shows the best and the second column the worst case. Green denotes true positive, blue false negative and red false positive segmented bounding box pixels.

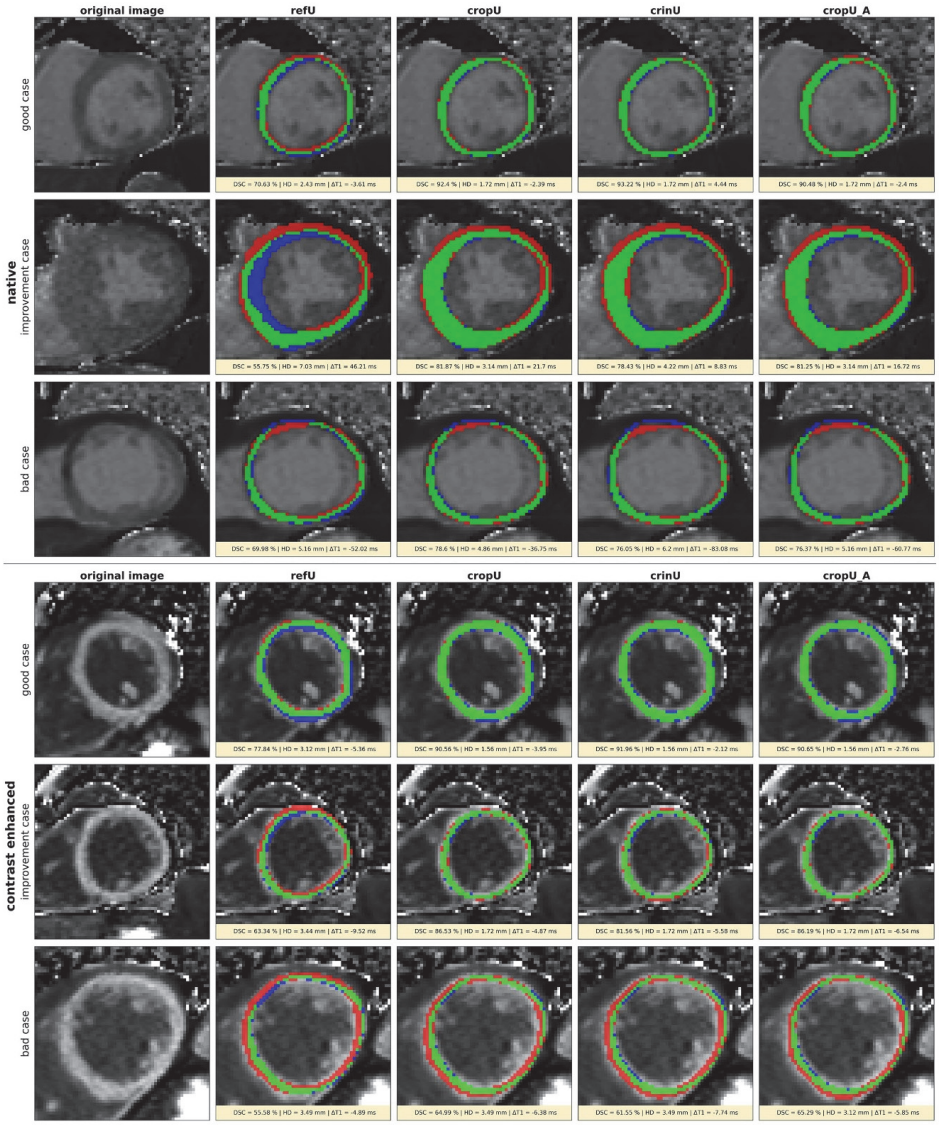

Figure 3. Example results of the automated segmentation in refU, cropU, crinU and cropU_A. The first column shows the original image, the second column the refU segmentation, the third column the cropU segmentation, the fourth column the crinU and the fifth column the cropU_A segmentation. The upper block corresponds to native and the lower block to contrast enhanced data; respectively in each block the first row shows a fairly good case across all four pipelines, the second row shows a case that is improved in cropU and crinU compared to refU and the third row shows a poor case across all four pipelines. Green denotes true positive, blue false negative and red false positive segmented pixels.

Metric	refU	cropU	crinU	cropU_A
Native				
Geometric				
DSC (%)	72.79 ± 8.08	81.06 ± 5.57*	81.22 ± 5.52*	81.13 ± 5.83*
HD (mm)	3.74 ± 1.37	2.95 ± 1.06*	3.01 ± 1.20*	2.98 ± 1.16*
Quantitative				
ME (ms)	− 7.22 ± 17.19	− 6.00 ± 14.67	− 5.24 ± 16.40*	− 3.88 ± 16.10*
MAE (ms)	14.22 ± 12.06	11.94 ± 10.43	12.45 ± 11.89	12.45 ± 11.89
RMSE (ms)	18.64	15.85	17.22	16.56
CI (ms)	− 11.38 / − 3.05	− 9.56 / − 2.44	− 9.22 / − 1.26	− 7.79 / 0.02
CV (%)	2.38	2.45	3.13	4.15
r (Pearson)	0.97*	0.97*	0.97*	0.97*
CoD (%)	94.09	94.09	94.09	94.09
τ (Kendall)	0.80*	0.83*	0.82*	0.83*
Contrast enhanced				
Geometric				
DSC (%)	71.41 ± 8.54	80.70 ± 10.31*	79.18 ± 10.20*	80.15 ± 10.21*
HD (mm)	3.83 ± 1.44	3.08 ± 1.72*	3.35 ± 1.90*	3.27 ± 2.05*
Quantitative				
ME (ms)	5.23 ± 8.14	4.45 ± 8.39	5.17 ± 7.27	4.57 ± 7.85
MAE (ms)	5.89 ± 7.67	5.32 ± 7.87	6.07 ± 6.54	5.07 ± 7.53
RMSE (ms)	9.67	9.50	8.92	9.08
CI (ms)	1.70 / 8.77	0.81 / 8.10	2.01 / 8.34	1.15 / 7.98
CV (%)	1.56	1.89	1.41	1.72
r (Pearson)	0.98*	0.98*	0.98*	0.98*
CoD (%)	96.04	96.04	96.04	96.04
τ (Kendall)	0.91*	0.91*	0.92*	0.94*

Table 2. Overview of the geometric and quantitative results for refU, cropU, crinU and cropU_A in the T1 map test dataset separated for native and contrast enhanced T1 maps. *DSC* dice similarity coefficient, *HD* Hausdorff distance, *ME* mean error, *MAE* mean absolute error, *RMSE* root-mean-squared error, *CI* confidence interval, *CV* coefficient of variation, *r* Pearson's correlation coefficient, *CoD* coefficient of determination, *τ* Kendall's Tau coefficient; values are given as mean ± standard deviation. *Denotes statistical significance with a significance level of $p < 0.05$.

The geometric quality improved significantly for all CASEG pipelines in comparison to the refU across the native and CE test datasets. The DSC improved significantly ($p < 0.05$) from around 72% towards 80% while the HDs were minimized significantly ($p < 0.05$) from above 3.70 mm to around 3.00 mm in all CASEG pipelines with the exception of a significant ($p < 0.05$) reduction to only 3.35 mm and 3.27 mm for CE data in crinU and cropU_A respectively. The corresponding boxplots in Fig. 4 illustrate these results and show that in some cases even a DSC of more than 90% was achieved in all three CASEG and across native and CE data, while refU reached a maximum DSC of 86.38% in native and 83.56% in CE data only. On the other hand, given 70% as threshold conventionally assumed as a good DSC[8], there were still cases left in cropU, crinU and cropU_A below that margin. The HD showed a minimization in the CASEG but remained with an average distance of 2.95 mm to 3.35 mm within the range of two to three pixels deviation.

Numerically ME, MAE and RMSE were consistently reduced in cropU, crinU and cropU_A compared to refU in the native as well as CE test data except for MAE in crinU for CE test data as shown in Table 2. Neither ME nor MAE showed a significant ($p < 0.05$) improvement compared to refU except for ME in crinU and cropU_A for native data. The CI of all pipelines stayed within the equivalence margin of 24.5 ms[11]. A visual support of this result is shown in the Supplemental Material S6. The CV of the CE test data stayed below the CV of the corresponding native test data in all pipelines. As a consequence, the CE results were assumed to remain in an adequate equivalence range based on CE data only.

Furthermore, all four pipelines showed a very strong linear correlation in native and CE test data, a strong monotonic correlation in the native test data and a very strong monotonic correlation in the CE data as shown in Table 2. The CoD was at least 94.09% such that the majority of the variation in the predicted average T1 time was explained by the variation of the targeted average left ventricular myocardial T1 time. Figure 5 shows the quantitative results as correlation- and Bland–Altman-plots in refU, cropU, crinU and cropU_A separately for native and CE test data. The plots indicate 20 native cases in refU, 11 native cases in cropU, 12 native cases in crinU, 11 native cases in cropU_A and 1 CE case in all pipelines that exceeded the limits of equivalence.

The Bland–Altman plots show that the limits of agreement differ only slightly between refU, cropU, crinU and cropU_A across the native as well as the CE test data. Further, the majority of those cases that exceeded the

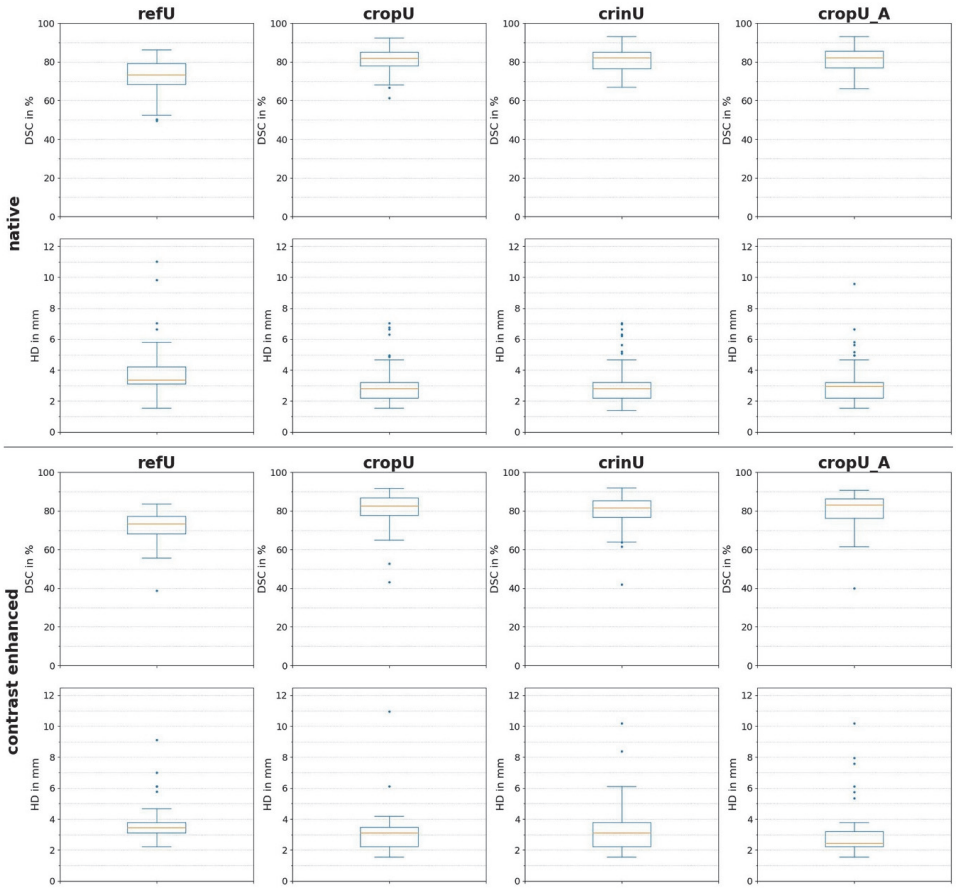

Figure 4. Geometric results of the automated segmentation. The first column shows the geometric results for refU, the second column for cropU, the third column for crinU and the fourth column for cropU_A. The upper block corresponds to native and the lower block to contrast enhanced data; respectively in each block the first row shows the boxplots of the Dice Similarity Coefficient (DSC) and the second row shows the boxplots of the Hausdorff Distance (HD).

equivalence margin were underestimating the expected average T1 time. This was confirmed in the histograms of disjoint pixel values as shown in Fig. 6. The histograms show that false negative segmented pixels in the native test dataset tend towards higher T1 values whereas in the CE test dataset towards lower T1 values independent of any of the three pipelines. Individual outliers in the native histograms are outside of the plotted range, but occur rarely on values above 2000 ms.

Finally, the coherence analysis in Fig. 6 shows the relationship between the DSC and the absolute T1 error. While in the native test data refU showed a weak and all CASEG showed a moderate linear correlation, the linear correlation in the CE test data was moderate in refU and strong in all CASEG. In contrast to that, the rank order stability was only weak across all test data and pipelines except for a moderate stability in the CE test data for crinU and cropU_A. Facing the maximum Pearson correlation coefficient of 0.81, the maximum CoD only reached a value of 65.61% implicating that more than a third of the variation is not explained. Nonetheless, except for the rank order stability correlation in CE data for refU, both correlation indices are significant ($p < 0.05$) in any pipeline. Further, it shows that most cases in CASEG and almost half of the cases in refU that exceeded the 24.5 ms equivalence margin had a DSC above 70% which is assumed with a good geometric result[8].

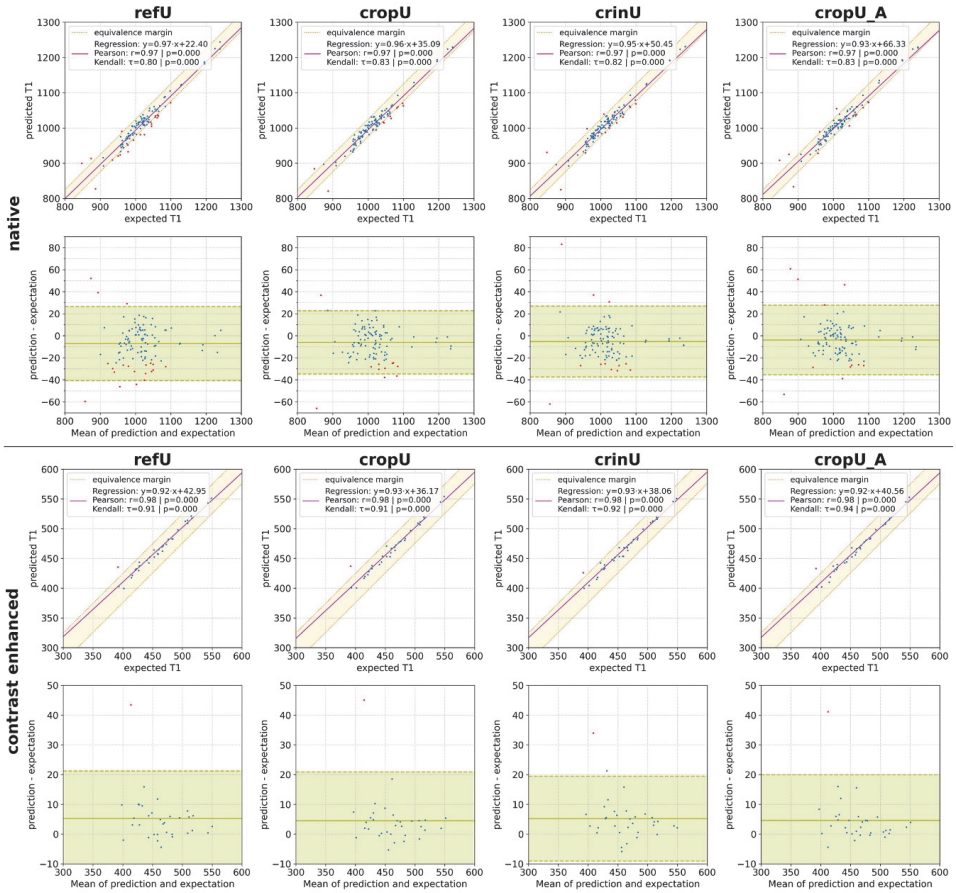

Figure 5. Quantitative results of the automated segmentation. The first column shows the quantitative results for refU, the second column for cropU, the third column for crinU and the fourth column for cropU_A. The upper block corresponds to native and the lower block to contrast enhanced data; respectively in each block the first row shows the correlation plot including the linear regression and the equivalence margin whereas the second row shows Bland–Altman-plots including the limits of agreement. Blue dots represent cases within the equivalence margin while red dots represent cases exceeding the equivalence margin.

Discussion

In CMR, the development of automated segmentation methods based on CNNs aims to substitute the necessity of an expert segmentation[30]. More complex network structures showed an improvement in segmentation quality[8,9,14] while it is also known that the segmentation quality highly depends on the input data quality[31–33]. Hence, this study explored the impact of an upstream object detection as a quality enhancement of input data on the segmentation quality of parametric T1 maps. Our main findings show a significantly improved segmentation in the geometric domain when using an ODA as a pre-processing step in a CASEG pipeline with a U-Net based segmentation CNN while in the quantitative domain a consistent but statistically not significant improvement in the estimation of the average T1 times was observed.

Dataset. The test dataset consisted of midventricular and basal slices only as those are recommended as stable slice location for a T1 map acquisition[4]. This differs from datasets described in the literature on automated segmentation methods for parametric T1 maps because either mid-ventricular slices only[9] or the whole short axis stack were used[8,10]. Apart from this, a comparison of our models with those of the literature is restricted

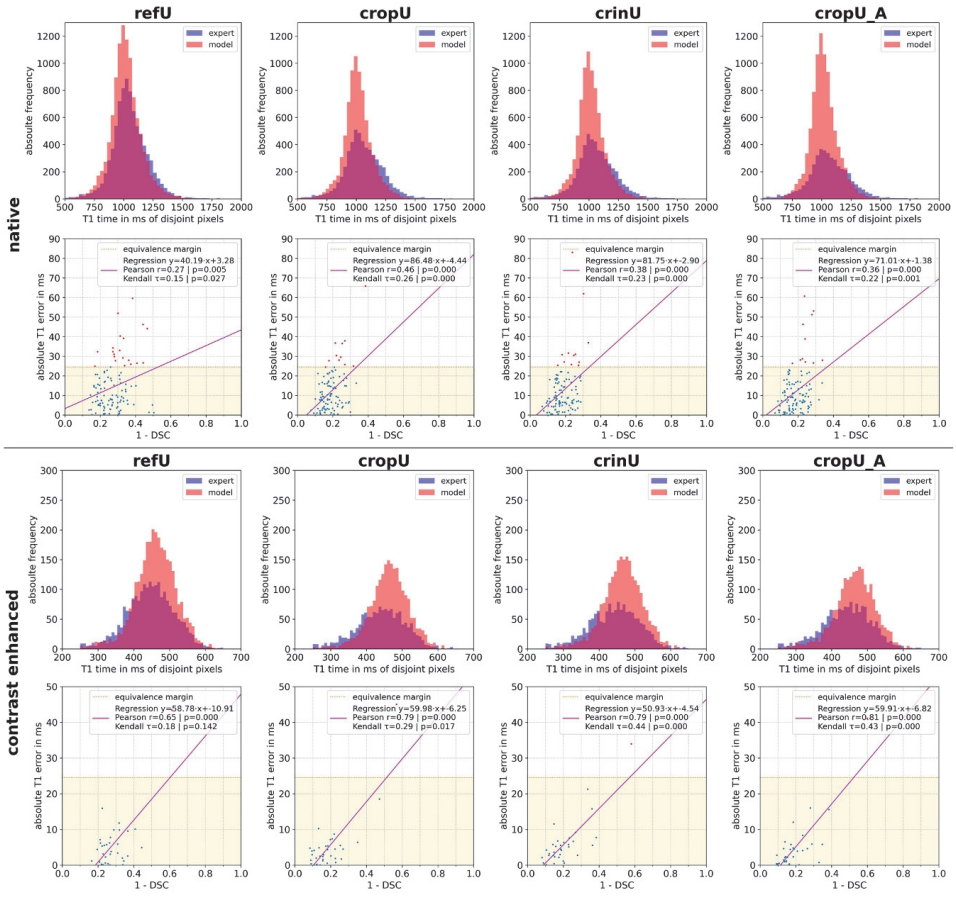

Figure 6. Coherence analysis of the automated segmentation. The first column shows the coherence analysis for refU, the second column for cropU, the third column for crinU and the fourth column for cropU_A. The upper block corresponds to native and the lower block to contrast enhanced data; respectively in each block the first row shows histograms of disjoint segmented pixel values of the expert ground truth and the pipeline model and the second row shows the correlation plot between Dice Similarity Coefficient (DSC) and the absolute T1 error including the linear regression. Blue dots represent cases within the equivalence margin while red dots represent cases exceeding the equivalence margin.

due to the lack of a common data basis in general. Furthermore, the ShMOLLI sequence as used in the literature gives different T1 mapping results than the used MOLLI sequence[34]. Therefore, a plausible comparison in the quantitative domain is limited.

ODA. Object detection is used for the semantic understanding and localisation of objects in images[15]. While the classical use-cases of ODAs are the detection of multiple objects from numerous possible categories in a single image leading to highly complex network structures[15], the ODA in the proposed CASEG pipelines had to find exactly one object from only one possible object class in an image. Hence, the use of a simple U-Net as ODA showed sufficiently good results for native and CE T1 maps in line with the results of Niu et al. who showed a DSC of 92.4 ± 3.6% in native CINE images for the left ventricular myocardial detection[16]. Our DSC results for the ODA were considerably above the 70% margin conventionally assumed as a good result[8].

The ODA was used as the first step in a CASEG pipeline to increase the ratio of relevant pixels by cropping the image to a ROI representing an enlarged BB section. The applied magnification factor of 1.50 corresponds

specifically to the used MOLLI T1 map short axis test dataset with our U-Net based ODA. As this factor is a freely adjustable hyperparameter in CASEG, an adaption is potentially necessary in other scenarios like different datasets or ODA networks. While the ODA network has a direct impact on the necessary magnification factor according to its prediction performance, two-dimensional quantitative data can be acquired in different orientations to meet the specific anatomy. Therefore, the optimal magnification factor is expected to be different in long axis views. Nonetheless, the stable DSC results across the magnification factor range of 1.30 to 2.50 for cropU and crinU showed similar performances even in out-of-optimum values for the magnification factor.

Although the ratio of relevant pixels was significantly ($p < 0.05$) increased in the ROI section compared to the original image, the result shows, that the majority of the pixels still belonged to background information. However, assuming a perfectly matched BB, this ratio could only be maximized to about twice the value gained, so that four out of five pixels would still belong to background information. The major reasons for this were the rather circular shape of the myocardium compared to the rectangular BB and the classification of the blood pool inside the myocardium as background information. The substantial variance in the ratio of relevant pixels within the 1.5 times enlarged BB as shown in the Supplemental Material S5 depended on the BB quality on the one hand and on the wall thickness of the myocardium on the other hand.

CASEG. Although a complex network structure, analogous to the DoubleU-Net by Jha et al.[35], could be used to integrate a whole CASEG pipeline into one network, the main idea of this work was to have a separated preprocessing step. Consequently, the ODA and the segmentation CNN in a CASEG pipeline are potentially interchangeable with other network structures. An alternative cropU structure with direct prediction of an enlarged BB omits the necessity of a magnification factor enlargement step at an equivalent outcome. Considering uncertainties during model training, cropU and cropU_A can be regarded as equivalent.

Our results emphasize that the upstream object detection improves the geometric segmentation quality in U-Net based automatic segmentation. Although the U-Net[12] is a common CNN for medical image segmentation, one could expect that novel CNN architectures enable potential further performance gains[8,9,14]. The classical U-Net, as our refU pipeline, has been used in prior studies as a benchmark CNN as well. While the basic structure of those U-Nets is similar, implementation details such as hyperparameter settings potentially differs from the original and definitely from our refU network[8–10] such that the performance comparability is limited. While Farrag et al. reached a DSC of 82.7% in native and 74.1% in CE T1 maps[10], Puyol-Antón et al. showed a DSC of 78%[9] and Hann et al. a DSC of 83.13%[8] in U-Net based segmentation of native T1 maps. While the refU is inferior to the classical U-Nets in those studies for the native dataset, both CASEG pipelines, cropU and crinU, were able to align with those results. For the CE dataset, refU performance was inferior to the results of Farrag et al. while cropU and crinU were outperforming it. However, none of the pipelines could reach geometric results of 84% as in the probabilistic hierarchical segmentation network[9] by Puyol-Antón et al. or 85% as in the quality control driven framework[8] by Hann et al. potentially due to their more complex segmentation network structure. With respect to an intra-observer performance of 72% DSC and 15.61 mm HD in native data and 83% DSC and 9.03 mm HD in CE data the CASEG pipelines showed a robust geometric outcome compared to a human reader[13]. However, errors made by the automated segmentation are prone to be atypical as compared to a human reader[30] such that the human segmentation is not necessarily substitutable by a completely unsupervised CASEG pipeline at the current stage.

Facing the quantitative domain of actual T1 values, no significant improvement in cropU and crinU compared to refU could be observed. This is at first glance counter-intuitive as a higher geometric accordance is assumed to coincide with a lower quantitative deviation. This was also shown in the coherence plot of Fig. 6 with a maximum CoD of 62.41% between DSC and the absolute T1 error underlining that an improved geometric result does not necessarily yield an improved quantitative result.

Taking into account, that the majority of the cases exceeding the equivalence margin in native T1 maps were underestimating the expert segmentation, the false negative segmented pixels belong to tissue that is assumed to contain blood. This agrees with the histograms in Fig. 6 and holds for the CE test data as well, as in CE blood has lower T1 values. Consequently, the border pixels are crucial as the impact of these disjoint pixels may be sufficient to impair improvements in the quantitative domain comparable to the significantly improved segmentation from a geometric point of view.

Comparing the quantitative results with literature values, the ME for the native dataset in refU, cropU and crinU lie in the published range of 4.6 ms[8], 8 ms[10] and 12.4 ms[9] while the ME for the CE test data were worse than the 2 ms in the proposed segmentation method by Farrag et al. but much better than the ME in their comparative U-Net model with 37 ms[10]. However, the MAEs in the native dataset were slightly exceeding in all pipelines the result of 11.3 ms by Hann et al.[8]. Nonetheless, the CI of the quantitative results stayed in all cases within the intra-observer equivalence margin[11].

As the segmentation quality depends on the input data quality[31–33], we were able to show that the ODA in a CASEG enhancing the input data quality results in geometric improvements. However, partial volume effects along the endocardial contour may have an important negative impact on the quantitative outcome. In contrast to this study, the software cvi42 internally provides the possibility to use a kind of safety margin by moving the contours towards the middle of the myocardium by a predefined amount in order to compensate to a certain degree false positive segmented pixels at both borders. However, this procedure is not a standard option in all commercially available postprocessing solutions.

Finally, the CASEG as well as all other automated segmentation models found in the literature work with loss functions solely based on geometric agreement[8–10] and neglect the quantitative domain. Therefore, the punishment for false positive segmented pixel during training of the models are equally independent of the actual T1 value. In conclusion it was shown in our study that an improved geometric congruence does not result in

a significant minimization of T1 value deviation. Nonetheless, a significant ($p < 0.05$) but mainly only weak to moderate correlation between geometric congruence and quantitative deviation were shown. Additionally, the proposed network by Hann et al. shows the highest geometric congruence at lowest T1 deviation which suggests a DSC cutoff margin somewhere between 80 and 85% where the influence of the disjoint segmented pixels attenuate due to the high geometric overlap.

Conclusion

The upstream object detection enables a significantly improved performance in the automated segmentation of parametric T1 maps from a geometric point of view compared to a standalone CNN. However, the quantitative measure could not be improved accordingly. Most likely the border pixels comprising partial volume effects between myocardium and blood play a key role in the discrepancy between geometric and quantitative results. As the quantitative domain is not represented in the training of the CNNs, segmentation of quantitative data like parametric T1 mapping may suffer from its absence. All in all, CASEG is well applicable for the improvement of segmentation tasks and this general approach may provide a viable extension to novel segmentation frameworks.

Outlook. While this study showed the potential of an ODA in an automated segmentation pipeline, a future step could be the exchange of the U-Net based segmentation CNN in the CASEG with a higher performing architecture[8,9] or more recent model adaptions[14]. This may provide an additional performance gain by the complex network structure in conjunction with the enhanced input data. Furthermore, it is worth investigating the combination of two CASEG pipelines for the segmentation of the myocardium by having one CASEG pipeline segmenting the blood pool area, which belongs to the endocardial contour, while the other CASEG pipeline segments the joined area of blood pool and myocardium, which represents the epicardial contour. The difference of both would return the myocardial segmentation and due to the hole-free segmentation masks, the ratio of relevant pixels may vastly increase. The latter enables the possibility of a better border definition.

Finally, one of the most crucial aspects in our study as well as in others is the definition of a loss function purely depending on the geometric concordance. A loss function taking both, the geometric as well as the quantitative deviation, into account would be preferable. As the domains are based on different physical units, the definition of such a loss function requires further investigation.

Limitation. The used dataset is composed of available segmented parametric T1 maps coming from different studies and scanners. A prior selection in order to have equipartition in specific characteristics of the used dataset was omitted. Consequently, the dataset consisted of fewer CE images than native images. Furthermore, the results are limited to midventricular and basal slices only in the test dataset.

The hyperparameter setup of the U-Nets was chosen as the best performing one across multiple tested scenarios. However, this setup might be optimized for our specific dataset and not generalize well to others. Additionally, not all possible hyperparameter setups were tested due to its endless combination possibilities.

The magnification factor of the ODA predicted BB was based on and optimized for our test dataset. An adapted evaluation of the factor is potentially necessary when using different datasets or CNN structures. This work used U-Net based CNN models as case study of CASEG. However, reachable performance gains might be different in other network structures such as TransUNet[14]. The applied equivalence margin is the strictest one as it is based on an intra-observer variability. As the equivalence margin was defined on native T1 map data only, its application on CE T1 maps is questionable. Consequently, an adequate equivalence margin for CE T1 maps or in its usage in a ECV map is currently missing and were substituted in this study by the analysis of the CV.

Data availability

The trained models and used dataset are available on request in an anonymized manner by contacting the first (DV: darian-steven.viezzer@charite.de) or last (JSM: jeanette.schulz-menger@charite.de) author. The source code can be accessed via GitHub under the URL: https://github.com/DSV-CUB/CASEG or in the Supplemental Material S7.

Received: 11 October 2022; Accepted: 27 January 2023
Published online: 06 February 2023

References

1. Guo, R. *et al.* Emerging techniques in cardiac magnetic resonance imaging. *J. Magn. Reson. Imaging* **55**, 1043–1059 (2022).
2. Puntmann, V. O. *et al.* Society for Cardiovascular Magnetic Resonance (SCMR) expert consensus for CMR imaging endpoints in clinical research: Part I: Analytical validation and clinical qualification. *J. Cardiovasc. Magn. Reson.* **20**, 67–67 (2018).
3. Kammerlander, A. A. *et al.* T1 mapping by CMR imaging: From histological validation to clinical implication. *JACC Cardiovasc. Imaging* **9**, 14–23 (2016).
4. Messroghli, D. R. *et al.* Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *J. Cardiovasc. Magn. Reson.* **19**, 75 (2017).
5. Čelutkienė, J. *et al.* Innovative imaging methods in heart failure: A shifting paradigm in cardiac assessment. Position statement on behalf of the Heart Failure Association of the European Society of Cardiology. *Eur. J. Heart Fail.* **20**, 1615–1633 (2018).
6. Taylor, A. J., Salerno, M., Dharmakumar, R. & Jerosch-Herold, M. T1 Mapping: Basic techniques and clinical applications. *JACC Cardiovasc. Imaging* **9**, 67–81 (2016).
7. Antonopoulos, A. S. *et al.* Machine learning of native T1 mapping radiomics for classification of hypertrophic cardiomyopathy phenotypes. *Sci. Rep.* **11**, 23596 (2021).
8. Hann, E. *et al.* Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Med. Image Anal.* **71**, 102029 (2021).

9. Puyol-Antón, E. *et al.* Automated quantification of myocardial tissue characteristics from native T1 mapping using neural networks with uncertainty-based quality-control. *J. Cardiovasc. Magn. Reson.* **22**, 60 (2020).
10. Farrag, N. A., Lochbihler, A., White, J. A. & Ukwatta, E. Evaluation of fully automated myocardial segmentation techniques in native and contrast-enhanced T1-mapping cardiovascular magnetic resonance images using fully convolutional neural networks. *Med. Phys.* **48**, 215–226 (2021).
11. Zange, L. *et al.* Quantification in cardiovascular magnetic resonance: Agreement of software from three different vendors on assessment of left ventricular function, 2D flow and parametric mapping. *J. Cardiovasc. Magn. Reson.* **21**, 12 (2019).
12. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv150504597 Cs* (2015).
13. Fadil, H. *et al.* A deep learning pipeline for automatic analysis of multi-scan cardiovascular magnetic resonance. *J. Cardiovasc. Magn. Reson.* **23**, 47 (2021).
14. Chen, J. *et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. Preprint at https://doi.org/10.48550/arXiv.2102.04306 (2021).
15. Zhao, Z.-Q., Zheng, P., Xu, S. & Wu, X. Object Detection with Deep Learning: A Review. *ArXiv180705511 Cs* (2019).
16. Niu, Y., Qin, L. & Wang, X. Myocardium detection by deep SSAE feature and within-class neighborhood preserved support vector classifier and regressor. *Sensors* **19**, 1766 (2019).
17. Muehlberg, F. *et al.* Native myocardial T1 time can predict development of subsequent anthracycline-induced cardiomyopathy. *ESC Heart Fail.* **5**, 620–629 (2018).
18. Birukov, A. *et al.* Myocardial evaluation of post-preeclamptic women by CMR: Is early risk stratification possible?. *JACC Cardiovasc. Imaging* **13**, 1291–1293 (2020).
19. Blaszczyk, E. *et al.* Progressive myocardial injury in myotonic dystrophy type II and facioscapulohumeral muscular dystrophy 1: A cardiovascular magnetic resonance follow-up study. *J. Cardiovasc. Magn. Reson.* **23**, 130 (2021).
20. Stengl, H. *et al.* Cardiomyocyte Injury following acute ischemic stroke: Protocol for a prospective observational cohort study. *JMIR Res. Protoc.* **10**, e24186 (2021).
21. Jadon, S. A survey of loss functions for semantic segmentation. In *2020 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB* 1–7. https://doi.org/10.1109/CIBCB48159.2020.9277638 (2020).
22. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* (2017).
23. Prechelt, L. Early stopping: But when? In *Neural Networks: Tricks of the Trade* 2nd edn (eds Montavon, G. *et al.*) 53–67 (Springer, 2012).
24. Shorten, C. & Khoshgoftaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 60 (2019).
25. Taylor, R. Interpretation of the correlation coefficient: A basic review. *J. Diagn. Med. Sonogr.* **6**, 35–39 (1990).
26. Martin, B. J. & Altman, D. G. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**, 307–310 (1986).
27. Abadi, M. *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv160304467 Cs* (2016).
28. Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
29. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
30. Hadler, T. *et al.* Introduction of Lazy Luna an automatic software-driven multilevel comparison of ventricular function quantification in cardiovascular magnetic resonance imaging. *Sci. Rep.* **12**, 6629 (2022).
31. Peng, P. *et al.* A review of heart chamber segmentation for structural and functional analysis using cardiac magnetic resonance imaging. *Magma N. Y. N* **29**, 155–195 (2016).
32. Luca, A. R. *et al.* Impact of quality, type and volume of data used by deep learning models in the analysis of medical images. *Inform. Med. Unlocked* **29**, 100911 (2022).
33. Dodge, S. & Karam, L. *Understanding How Image Quality Affects Deep Neural Networks*. Preprint at http://arxiv.org/abs/1604.04004 (2016).
34. Heidenreich, J. F. *et al.* T1- and ECV-mapping in clinical routine at 3 T: Differences between MOLLI, ShMOLLI and SASHA. *BMC Med. Imaging* **19**, 1–9 (2019).
35. Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P. & Johansen, H. D. *DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation*. https://doi.org/10.48550/arXiv.2006.04868 (2020).

Acknowledgements
This study was supported by the BMBF (Bundesministerium für Bildung und Forschung)/DZHK (German Centre for Cardiovascular Research) via project FKZ81Z0100208 and complies with the declaration of Helsinki. We thank Yashraj Bhoyroo, Jan Gröschel and Johanna Kuhnt for contributing their segmentations in those dataset cases that belong to their research studies.

Author contributions
D.V., T.H., C.A., J.W. and S.L. designed and implemented the methods. D.V., E.B., M.F. and J.S.M. collected the dataset. D.V., E.B., M.F., T.H.G., S.L. and J.S.M. provided statistical analysis and interpretation of the data. S.L. and J.S.M. supervised the whole work. D.V. drafted the manuscript and all authors contributed in a critical revision of the manuscript. The final version is approved by all authors.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-28975-5.

Correspondence and requests for materials should be addressed to J.S.-M.

Articles

Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept

Darian Viezzer,[a,b,c,*] Thomas Hadler,[a,b,c] Jan Gröschel,[a,b,c] Clemens Ammann,[a,b,c] Edyta Blaszczyk,[a,b] Christoph Kolbitsch,[d] Simone Hufnagel,[d] Riccardo Kranzusch-Groß,[e] Steffen Lange,[f] and Jeanette Schulz-Menger[a,b,c,g]

[a]Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, ECRC Experimental and Clinical Research Center, Lindenberger Weg 80, 13125 Berlin, Germany
[b]Working Group on Cardiovascular Magnetic Resonance, Experimental and Clinical Research Center, A Joint Cooperation Between the Charité – Universitätsmedizin Berlin and the Max-Delbrück-Center for Molecular Medicine, Berlin, Germany
[c]DZHK (German Centre for Cardiovascular Research), Partner Site Berlin, Berlin, Germany
[d]Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany
[e]Universitätsklinikum Schleswig-Holstein, Klinik für Radiologie und Nuklearmedizin, Lübeck, Germany
[f]Hochschule Darmstadt (University of Applied Sciences), Faculty for Computer Sciences, Darmstadt, Germany
[g]Helios Hospital Berlin-Buch, Department of Cardiology and Nephrology, Berlin, Germany

eBioMedicine
2024;102: 105055
Published Online 14 March 2024
https://doi.org/10.1016/j.ebiom.2024.105055

Summary

Background In cardiovascular magnetic resonance imaging parametric T1 mapping lacks universally valid reference values. This limits its extensive use in the clinical routine. The aim of this work was the introduction of our self-developed Magnetic Resonance Imaging Software for Standardization (MARISSA) as a post-hoc standardisation approach.

Methods Our standardisation approach minimises the bias of confounding parameters (CPs) on the base of regression models. 214 healthy subjects with 814 parametric T1 maps were used for training those models on the CPs: age, gender, scanner and sequence. The training dataset included both sex, eleven different scanners and eight different sequences. The regression model type and four other adjustable standardisation parameters were optimised among 240 tested settings to achieve the lowest coefficient of variation, as measure for the inter-subject variability, in the mean T1 value across the healthy test datasets (HTE, N = 40, 156 T1 maps). The HTE were then compared to 135 patients with left ventricular hypertrophy including hypertrophic cardiomyopathy (HCM, N = 112, 121 T1 maps) and amyloidosis (AMY, N = 24, 24 T1 maps) after applying the best performing standardisation pipeline (BPSP) to evaluate the diagnostic accuracy.

Findings The BPSP reduced the COV of the HTE from 12.47% to 5.81%. Sensitivity and specificity reached 95.83% / 91.67% between HTE and AMY, 71.90% / 72.44% between HTE and HCM, and 87.50% / 98.35% between HCM and AMY.

Interpretation Regarding the BPSP, MARISSA enabled the comparability of T1 maps independently of CPs while keeping the discrimination of healthy and patient groups as found in literature.

Funding This study was supported by the BMBF / DZHK.

Keywords: Cardiovascular Magnetic Resonance; T1 Mapping; Standardisation; Healthy Volunteers; Left Ventricular Hypertrophy; Amyloidosis

*Corresponding author. Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, ECRC Experimental and Clinical Research Center, Lindenberger Weg 80, 13125, Berlin, Germany.
E-mail address: darian-steven.viezzer@charite.de (D. Viezzer).

Articles

Research in context

Evidence before this study

Although parametric T1 mapping is considered as an important method in cardiovascular magnetic resonance (CMR), the lack of universal valid reference values has been mentioned in many studies as an obstacle to fully utilise T1 mapping in different cohorts, multi-side studies or current guidelines. Hence, literature research covering the time frame from September 2016 until October 2023 in PubMed and Google Scholar for articles on T1 mapping in CMR revealed only the z-Score approach as a way to define comparable values. However, the necessity of a healthy cohort examination whenever a technical change is performed, high volatility in the standard deviation of healthy volunteer examination and lack of accessibility seems to be obstacles of this approach.

Added value of this study

Our work introduces the Magnetic Resonance Imaging Software for Standardization (MARISSA) as an approach for the post-hoc standardisation of parametric T1 maps in CMR. As this standardisation pipeline can capture different settings, we analysed 240 different settings in a two-step approach and evaluated the best performing one. We were able to show that the choice of the pipeline setting is crucial for the success of the standardisation and that this proof-of-concept including the parameters age, sex, scanner and sequence is already good enough to differentiate healthy volunteers from patients with hypertrophic cardiomyopathy and amyloidosis as in a highly controlled intra-scanner-intra-sequence setting.

Implications of all the available evidence

Our results demonstrate that a post-hoc standardisation of parametric T1 maps is feasible. The implementation as a python software with a graphical user interface makes the standardisation procedure directly available and shareable on any common operating system. Although the described approach with four considered confounding parameters already allowed a comparable discrimination of two cardiovascular diseases from a healthy cohort, this work is a proof-of-concept that needs further investigations on more scanners, sequences and diseases but also other confounding parameters. Compared to the z-score, our standardisation pipeline does not require a re-examination of a healthy reference cohort on each site whenever a technical change occurs. This reduces the effort and costs to increasingly enable parametric T1 mapping. Consequently, this work is a further step forward to strengthen the establishment of parametric T1 mapping in the clinical routine, which in turn helps to improve the detection of cardiovascular diseases.

Introduction

Cardiovascular magnetic resonance (CMR) is recommended as the non-invasive imaging modality of choice for myocardial tissue characterisation in cardiovascular diseases (CVD).[1–3] This characterisation is enabled by quantitative methods such as parametric T1 mapping.[4,5] Its integration into clinical CMR routine protocol recently showed improved diagnostic accuracy for the detection of CVDs.[6] Although parametric T1 mapping is already increasingly used in clinical routine and turned from a research to a product sequence, reproducibility is limited to intra-institutional reference values.[4,7] This lack of universally applicable reference values is a major obstacle for a stronger assertiveness of parametric T1 mapping in the clinical routine and is caused by subject specific, technological and post-processing procedure variations.[7–9] These variations act as confounding parameters (CPs) on parametric T1 maps and thereby on the quantitative outcome, which in turn potentially influence the treatment of CVDs. Hence, every change in hard- or software may require new local reference values and thus a re-examination of a healthy reference cohort.

Recently, the reproducibility of parametric T1 mapping was validated across different scanners if CPs such as manufacturer, field strength, acquisition schemes and post-processing were kept constant.[10] However, the technical setups across institutions are manifold and cannot be globally aligned by force. As the influence of CPs such as age and sex,[11,12] sequence variants[13,14] or scanner models[12] on parametric T1 mapping were described in the literature, universal valid reference values are required to consider and consequently to minimise the induced CP's bias.

All in all, there is a gap of defining universal valid reference values in parametric T1 mapping based on a generalised approach. For that reason, the aim of this work is to introduce a generic post-hoc standardisation pipeline that enables comparability while maintaining diagnostic accuracy and reducing the amount of necessary healthy volunteer examinations. We propose that standardisation is enabled by estimating the impact of a CP relatively to a reference CP value. Consequently, parametric T1 mapping values are post-hoc transformable into values of a reference CP environment. This proposed transformation is embedded in the self-developed open-source Magnetic Resonance Imaging Software for Standardization (MARISSA) that is made available with this work.

Methods

In this work the four CPs: age,[11,12] sex,[11,12] scanner[12] and sequence variant[13,14] were chosen from literature to show a proof-of-concept for the proposed post-hoc standardisation of parametric T1 maps. This includes

the introduction of MARISSA as a software tool to setup those post-hoc standardisation pipelines and to demonstrate the diagnostic quality after standardisation. Therefore, the used data collection included three cohorts: healthy volunteers (Healthy), patients with left ventricular hypertrophy (LVH) including hypertrophic cardiomyopathy (HCM) and patients with amyloidosis (AMY). Both patient cohorts were included on account of a statistical significant differentiation from healthy volunteers in native T1 mapping with considerably higher T1 values in AMY and on average higher but partly overlapping ranges in HCM.[15] The following first sub-section Dataset describes the used data collection of the three included cohorts. A part of Healthy (Healthy train datasets/HTR) was used in a first step to estimate the individual CP impact. As different strategies exist to estimate the CP induced bias, the Confounding Parameters Impact Estimation (CPIE) sub-section covers detailed information about the different estimation strategies. The following sub-section on the Best Performing Standardisation Pipeline (BPSP), describes the evaluation of the best performing CPIE among all tested strategies with respect to the remaining healthy volunteers (Healthy test datasets/HTE). The before last subsection on Diagnostic Implication (DI) covers the evaluation of the diagnostic accuracy and intra-subject differences after applying the BPSP on the said HTE as well as the HCM and AMY cohorts and includes the Statistics. Finally, the Implementation sub-section describes briefly the MARISSA structure followed by the Ethical approval and Role of the funding source.

Dataset

The included retrospective data collection of the three cohort groups: Healthy, HCM and AMY consisted of midventricular slices only and originated from previous and ongoing studies of our working group or in which our working group participated until June 2023 considering scanners that are part of the Berlin CMR research network.[10] Age was the only numerical CP; all other considered CPs were categorical. The various origins in the data collection enabled variability in the concerned CPs while some individuals received multiple measurements, i.e. different sequences and/or scanners. Fig. 1 shows an overview of the total numbers in each cohort as well as the variation in the four concerned CPs. A detailed breakdown of the underlying data for each scanner-sequence combination for both sexes is provided in the Supplemental Material S1.

The segmentation of all parametric T1 maps was performed automatically with a subsequent visual inspection by two experts (JG and EB). First, all T1 maps were segmented using the cropU cascaded model[16] and a research deep learning segmentation model provided by Siemens Healthcare (version 21 hotfix, Siemens Healthcare GmbH, Erlangen, Germany). The expert chose the best segmentation out of either model or the intersection of both. If none was considered accurate enough by the experts, the segmentation was performed manually in cvi42 (version 5.13.7, Circle Cardiovascular Imaging, Calgary, Canada) as previously reported.[10]

Confounding Parameters Impact Estimation (CPIE)

The CPIE is a regression model training based on the HTR and represents the central part of the proposed standardisation pipeline. For that reason, the Healthy data collection were split into 85% training and 15% testing with respect to the number of subjects per study. Thus the HTR consisted of 214 subjects (814 T1 maps, 100 males / 114 females and 38.46 ± 15.20 years) while the HTE consisted of 40 subjects (156 T1 maps, 18 males / 22 females, 39.50 ± 15.89 years).

The CP impact can be solely estimated relatively to a reference as the absolute true T1 mapping value is unknown due to an intrinsic lack of accuracy or precision in T1 acquisition methods.[9] Therefore, we propose to define for each CP a reference value that is assumed with no bias. In this work, the reference CP values were set to 18 years, male, 3.0T Siemens Verio [syngo MR B17] and T1 Map MOLLI 5(3)3 b for the concerned CPs age, sex, scanner and sequence variant respectively.

Consequently, the regression estimates the difference between the examined apparent T1 mapping value in the according CP environment and the target T1 mapping value in the reference CP environment. The target T1 mapping value is defined as the mean T1 value of HTR subjects whose concerned CP value matches the reference CP value. During the fitting of the regression the difference between apparent and target T1 mapping value represents the dependent and the CP value(s) represent the independent variable(s). Whether one or multiple CPs are taken into account depends on the strategy setting. Table 1 lists an overview of possible settings for the standardisation pipeline that are explained in the following.

Each CPIE model is described by the regression-type, y-type, mode, bins and cluster-type. The linear regression is the most basic regression model used in a variety of disciplines[17] while the linear support vector regression[18] (LSVR) was implemented as an alternative linear model. Additionally, the random-forest[19] regression (RFR) and extra-trees[20] regression (ETR) were implemented as those are assumed to handle non-linear relationships better than linear models. All regression models were based on the scikit-learn package[21] and setup in the default setting, except for RFR and ETR, where the number of estimator trees were increased to 1000. While an increased number of trees enables an improved performance, an overfitting is excluded by its intrinsic structure.[19]

The chosen regression-type can either estimate the difference between apparent and target T1 value in absolute (ms) or relative (%) values according to the y-type setting. While the absolute case shifts all values equally

Articles

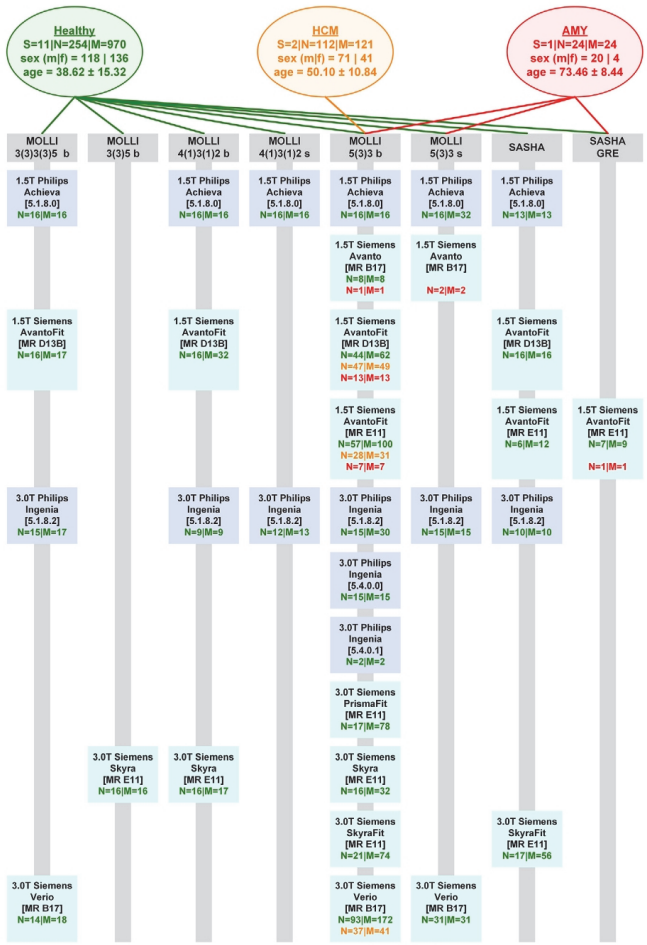

Fig. 1: Dataset Overview – The general overview of the Healthy (green), patients with left ventricular hypertrophy including hypertrophic cardiomyopathy (HCM, orange) and patients with amyloidosis (AMY, red) cohorts are given with S: number of studies, N: number of subjects, M: number of midventricular native parametric T1 maps, m: males and f: females. The age is given as mean ± standard deviation. The grey bars represent respective parametric T1 mapping sequences and the turquoise and blue boxes represent a scanner. For each scanner-sequence combination the number of subjects and respective T1 maps is itemised.

such that the intrinsic value spread maintains, the relative case weights the shift according to the absolute T1 value.

The mode, finally, defines whether a regression is performed for each CP individually (individual and cascaded) or all at once (ensemble). The individual mode fits a regression for each CP individually and independent. Any other CPs are kept constant to minimise their impact on the regression. Consequently, only a small portion of the HTR can be used for each CP regression fit. The cascaded mode works the same for the first CP regression but iteratively standardise the

Setting	Options
regression-type	extra-trees (ETR) linear linear support vector (LSVR) random-forest (RFR)
y-type	absolute relative
mode	cascaded ensemble individual
bins	1 to minimum number of segmented pixels
cluster-type	agglomerative average agglomerative complete agglomerative single agglomerative ward equal distant equal size gaussian mixture k-means

Table 1: **Standardisation pipeline setting – Each pipeline setting can capture exact one of the given possible values.**

HTR data, such that subsequent CPs must not consider previously fitted CPs. Therefore, the amount of useable data increases for later CPs. It is important to note that the outcome depends on the order of the considered CPs. Finally, the ensemble mode takes all CPs at once. The ensemble mode is expected to handle cross-dependent CPs best. Categorical variables are converted into category numbers in the ensemble mode, whereas in individual and cascaded mode each category receives its own regression model.

While regression-type, y-type and mode directly affects the regression model, the setting of a bin larger than one accounts for cross-dependencies between the CP value and the apparent T1 value. It is currently unknown, if higher T1 values are differently affected by a CP than lower ones. The setting of a bin larger than one requires the choice of a clustering algorithm that clusters the T1 values into bins. The different agglomerative clustering algorithms, the Gaussian mixture and k-means clustering were taken from the scikit-learn package[21] while equal distant and equal size clustering were self-implemented. For equal distant clustering all bins have the same width while for equal size clustering the T1 values are sorted and the same number of T1 values are used for each bin. Although the number of bins is mainly limited by the smallest possible number of segmented pixels, which would represent each pixel as its own cluster, we recommend ten bins or fewer as otherwise the number of T1 values in each bin is too small to be representative.

In the light of the standardisation pipeline, CPIE outputs the estimated bias and is therefore an integrative part of the pipeline. First, a T1 mapping dataset in Digital Imaging and Communications in Medicine (DICOM) format with a corresponding segmentation mask is expected as input. After extracting the segmented apparent T1 values, the CPIE is applied to estimate the CP induced bias. This bias is then subtracted from the apparent T1 values and the resulting standardised T1 values are returned as output.

Best performing standardisation pipeline (BPSP)

As the optimal standardisation pipeline setting is unknown, CPIE were evaluated for 240 different settings in two steps. First, bins were set to one, hence, cluster-type had no impact and all 24 combinations of regression-type, y-type and mode were fitted. The resulting standardisation pipelines were evaluated with respect to the coefficient of variation (COV, Equation (1)) of the mean T1 time in the HTE.

$$COV = \frac{\sigma}{\mu} \qquad \text{(Equation 1)}$$

The lower the COV, the less variability across the subjects of the HTE exist and, consequently, a better pipeline performance can be assumed. Considering the top three performing pipelines of this first step, fitting was performed for all combinations of two to ten bins and clustering algorithms resulting in additional 216 pipelines. The BPSP was evaluated as the one out of the 240 fitted standardisation pipelines with the lowest COV in the HTE group.

Diagnostic Implication (DI)

While the evaluation of the BPSP accounted for a minimisation of the inter-healthy-subject variability, the DI step assessed the discriminability between the healthy test cohort and patients. Therefore, the HTE, HCM and AMY cohort were standardised with the BPSP. A progression plot was used to show the value progression from before to after standardisation.

Statistics

All available retrospective datasets were included that were diagnosed as either Healthy, HCM or AMY. Age and sex were self-reported and all subjects were 18 years or older. Non-midventricular and contrast enhanced parametric T1 maps were excluded. The data were randomized, blinded and checked for artifacts by two experts (JG and EB).

The outcome statistics of the post-hoc standardisation is integrated in the DI. This includes the boxplots for each cohort before and after standardisation indicating the respective value spread. Further, confidence intervals (CIs) were calculated and statistics between the cohorts after standardisation were tested with an independent t-test and ANOVA test if all cohorts were normal distributed according to the Shapiro-Wilk-test, otherwise with the Mann-Whitney-U and Kruskal–Wallis test. Significance was assumed if both tests had a significance level of $\alpha \leq 0.05$. Furthermore, a receiver

Articles

operating characteristics (ROC) analysis was performed to evaluate the optimal threshold between the cohorts. The post-hoc standardisation ROC were compared to the intra-scanner-intra-sequence ROC before standardisation to evaluate the maintenance of the diagnostic accuracy. Further, evidence was assumed if the sum of sensitivity and specificity reached 150% or above as recommended in literature.[22] Finally, intra-subject progression plots were performed with HTE subjects that were measured in different CP environments.

Implementation

The core of MARISSA is a SQLite database in the backend, referred in the following as MARISSA DB, and an overlaying graphical user interface (GUI). Fig. 2 shows abstractly the structure and user interaction in MARISSA. The software was fully implemented in Python (Version 3.8, Python Software Foundation, Beaverton, USA) and is available in the Supplemental Material S2.[23] All necessary site packages, installation instructions and further detailed information are listed in the MARISSA User Manual in the Supplemental Material S3.

As MARISSA is not only implemented for parametric T1 mapping in CMR, the software works in separate projects with individual MARISSA DBs, that are exportable with or without data. Within MARISSA DB, the related tables are separated: On the one hand the active site, where the user manipulates data, settings and parameters and a passive site that contains all trained standardisation pipelines including a copy of all necessary information to reconstruct the training. This separation assures for a retraceable standardisation pipeline training whereas an export of the project without data loses the traceability while still maintaining the standardisation functionality.

The definition of CPs is based on DICOM tags. Standard DICOM tags are already available in MARISSA, while specific CPs like the sequence variant are extracted by string processing of the series description. The choice of the value representation defines the CP as either a numerical or categorical parameter. MARISSA supports also multi-value DICOM tags as long as the multiplicity remains stable. In this case, each value dimension is considered individually as an own CP. The DICOM standard gives more information about the DICOM tag composition.[24]

The import of DICOM data and segmentations includes a customisable description in order to enable cohort differentiation within the MARISSA DB. While training data must be imported, the standardisation pipeline is applicable on imported as well as external data. Applying the standardisation on a dataset exports the original DICOM data, an Excel table, a MARISSADATA file and a progression plot. The Excel table contains information about the CP values and the transformation of the segmented T1 values while the MARISSADATA file contains the same information as a pickled Python dictionary such that it can be imported and further processed in other Python applications. More detailed information about the usage of MARISSA is provided in the Supplemental Material S3.

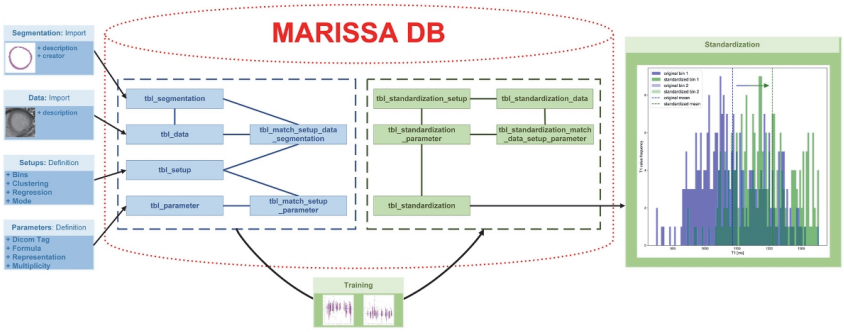

Fig. 2: MARISSA structure – The central element is the MARISSA DB based on a SQLite database with relational connected tables: The blue tables tbl_segmentation, tbl_data, tbl_setup, tbl_parameter, tbl_match_setup_data_segmentation and tbl_match_setup_parameter are those, where user interaction takes place by adding, editing or deleting information and data. When starting the training of a standardisation pipeline, the confounding parameter impact estimation is fitted. In order to track back the training, the necessary information is copied into the separate green tables tbl_standardization_setup, tbl_standardization_data, tbl_standardization_parameter and tbl_standardization_match_data_setup_parameter while the tbl_standardization stores the fitted regression models for the confounding parameter impact estimation. Applying the standardisation on a dataset will transform the T1 values into a range that represents the reference confounding parameter environment as depicted on the right site.

Ethical approval

This study was approved by the local ethics committee of the Charité Universitätsmedizin Berlin as retrospective study (study ID: EA 1253 21) and complies with the declaration of Helsinki. The requirement for written informed consent was acquired during the original clinical studies and was therefore waived in this study due to its retrospective design as approved by the local ethics committee of the Charité Universitätsmedizin Berlin (study ID: EA 1253 21). Due to institutional law, datasets cannot be shared.

Role of the funding source

This study was supported by the BMBF (Bundesministerium für Bildung und Forschung)/DZHK (German Centre for Cardiovascular Research) via project FKZ81Z0100208. The BMBF/DZHK had no influence on the design, execution or evaluation of this study.

Results

The results are given for the three steps Confounding Parameters Impact Estimation (CPIE), Best Performing Standardisation Pipeline (BPSP) and Diagnostic Implication (DI) separately as well as the Implementation of MARISSA.

Confounding Parameters Impact Estimation (CPIE)

The CPIE could be successfully trained without abortion on all 24 settings without and 216 settings with clustering. However, as the individual mode requires constant CP values for all CPs except the estimating one, the training with the used HTR could not include the sequences T1 map MOLLI 3(3)5 b and T1 map SASHA GRE due to variations in the other CPs. Consequently, these two sequences could not be standardised and acted like no bias.

Best performing standardisation pipeline (BPSP)

The top three settings across the 24 settings without clustering were LSVR regression on relative values in cascaded mode, ETR on relative values in ensemble mode and ETR on absolute values in ensemble mode with a COV of 5.98%, 6.10% and 6.23% respectively for the mean T1 value of the respective standardised HTE. Among all 240 trained pipelines, the BPSP was obtained with the LSVR regression on relative values in the cascaded mode with two bins and the agglomerative single clustering resulting in a COV of 5.81%.

Fig. 3 plots the COV for each trained standardisation pipeline including the best obtained COV and the COV of the unstandardised HTE of 12.47% revealing that some standardisation pipeline settings even worsen the uncertainty in the HTE.

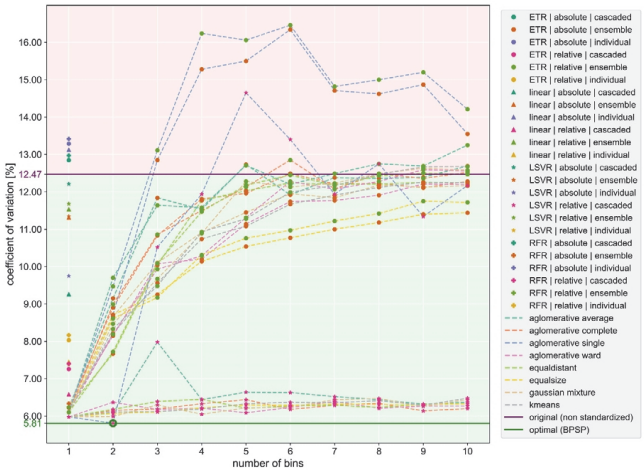

Fig. 3: Coefficient of variation (COV) in the 240 trained standardisation pipelines – The COV is plotted against the standardisation pipeline setting denoted as the number of bins on the x-axis, the regression-type by the scatter point marker style, the y-type and mode according to the scatter point colour and the clustering algorithm according to the dotted line for bins greater than one. The purple solid line represents the COV threshold of the unstandardised data with everything above in the red area means a worsening while everything below in the green area means an improvement of the intra-healthy-subjects variation. The green line with the inline circle shows the optimal COV reached among the 240 pipelines representing the best performing standardisation pipeline (BPSP).

Diagnostic Implication (DI)

Applying the BPSB on the three test cohorts HTE, HCM and AMY results in a COV of 5.81%, 4.46% and 6.05% respectively compared to an unstandardised COV of 12.47%, 9.56% and 6.06%. Hence, COVs for HTE and HCM showed an improvement after standardisation while the COV for AMY remained almost equal. Fig. 4 gives an overview of the respective cohort data before and after standardisation with the BPSP and the individual impact of each CP on every single test dataset.

Already before standardisation the HTE and HCM as well as HCM and AMY but not HTE and AMY were statistically significant different. However, the 25%–75% quantile ranges of HCM (999.57–1186.00 ms) and AMY (1087.84–1148.40 ms) were almost completely within the range of HTE (1007.81–1213.34) and the mean value of both patient cohorts (HCM: 1076.69 ms / AMY: 1118.39 ms) were lower than in HTE (1140.20 ms) while both patient groups are expected with significant higher T1 values.[15] This contradiction to literature were due to the mixture among CPs in the data collection, especially in regard of the occurring scanner and sequence combinations according to the detailed cohort dataset breakdown in the Supplemental Material S1. After standardisation, all cohorts revealed statistically significant difference from each other and no overlapping CIs. None of the cohorts, neither before nor after standardisation, could remain in the margin of the CI of the unstandardised HTR data, which captured the reference CP environment. However, after standardisation the HTE was closest to fit in whereas HCM and AMY were clearly above. The resulting T1 value ranges (mean ± standard deviation) after standardisation were 1136.78 ± 66.09 ms, 1186.27 ± 52.93 ms and 1337.62 ± 80.92 ms for HTE, HCM and AMY respectively.

The ROC analysis, as shown in Fig. 5, revealed an optimal threshold (sensitivity / specificity) of 1163.89 ms (71.90% / 72.44%) between HTE and HCM, 1204.46 ms (95.83% / 91.67%) between HTE and AMY and 1287.89 ms (87.50% / 98.35%) between HCM and AMY after standardisation. The differentiation between HTE and HCM were slightly below the 150% threshold for the sum of sensitivity and specificity while HTE and AMY as well as HCM and AMY were above it.[22] In all three post-hoc standardised ROC analysis sensitivity and specificity were in the range of unstandardised intra-scanner-intra-sequence differentiability. The sum of sensitivity and specificity increased or at least remained after standardisation within the intra-scanner-intra-sequence datasets although individual sensitivity and specificity values changed.

Finally, eight subjects of the HTE group received at least two different acquisitions. The intra-subject progression plot for each of these subjects is shown in Fig. 6. The plot shows a minimisation of the value spread after standardisation in all subjects but one. Nonetheless, all subjects showed a minimisation of the COV, which reflects a concentration of the acquisitions

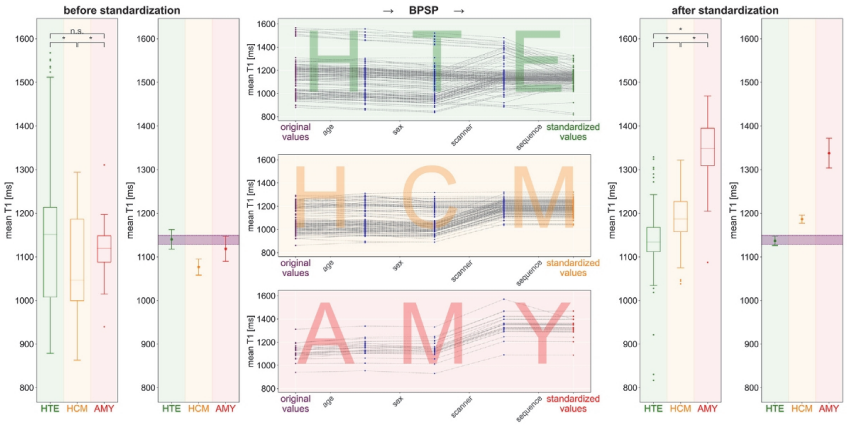

Fig. 4: Inter-cohort progression plot – On the left and right are respectively before and after standardisation with the best performing standardisation pipeline (BPSP) the boxplot of the mean T1 value for each cohort: Healthy test datasets (HTE), patients with left ventricular hypertrophy including hypertrophic cardiomyopathy (HCM) and patients with amyloidosis (AMY). The * denotes statistically significant differences and n. s. means not significant. Further, the confidence intervals are plotted against the purple area that represents the confidence interval of the unstandardised healthy data which captures the reference confounding parameter environment. In the middle the progression from original towards standardised values with the BPSP are plotted with detailed impact for each confounding parameter.

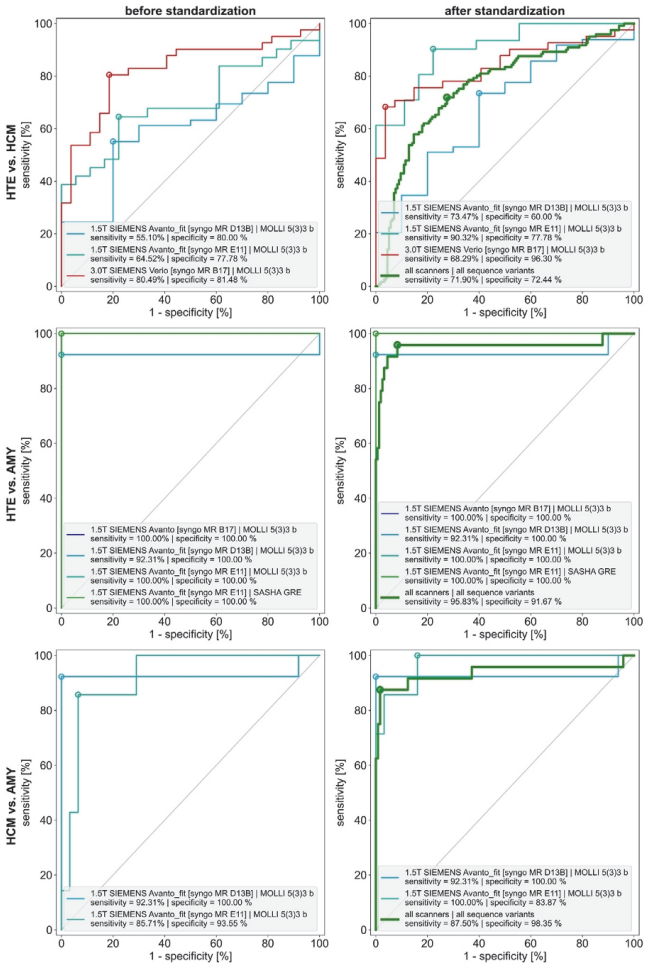

Fig. 5: ROC analysis – The left site represents before and the right site after standardisation. Each curve represents intra-scanner-intra-sequence data. After standardisation shows additionally the ROC analysis in a bold green curve that reflects the differentiability among all healthy test datasets (HTE), patients with left ventricular hypertrophy including hypertrophic cardiomyopathy (HCM) and patients with amyloidosis (AMY).

after standardisation. However, individual measurements, especially SASHA based parametric T1 maps, revealed high imprecision after standardisation.

Implementation

The MARISSA was successfully implemented as a Python software tool with a SQLite database backend. The overlaying GUI allows usability for programmers and clinicians alike. The definition of custom CPs, however, requires at least experience in the DICOM tag standard as well as Python string processing. Furthermore, data filtering in the GUI works by SQL commands via respective input fields. The software enables to work in projects that can be exported with and without data.

Articles

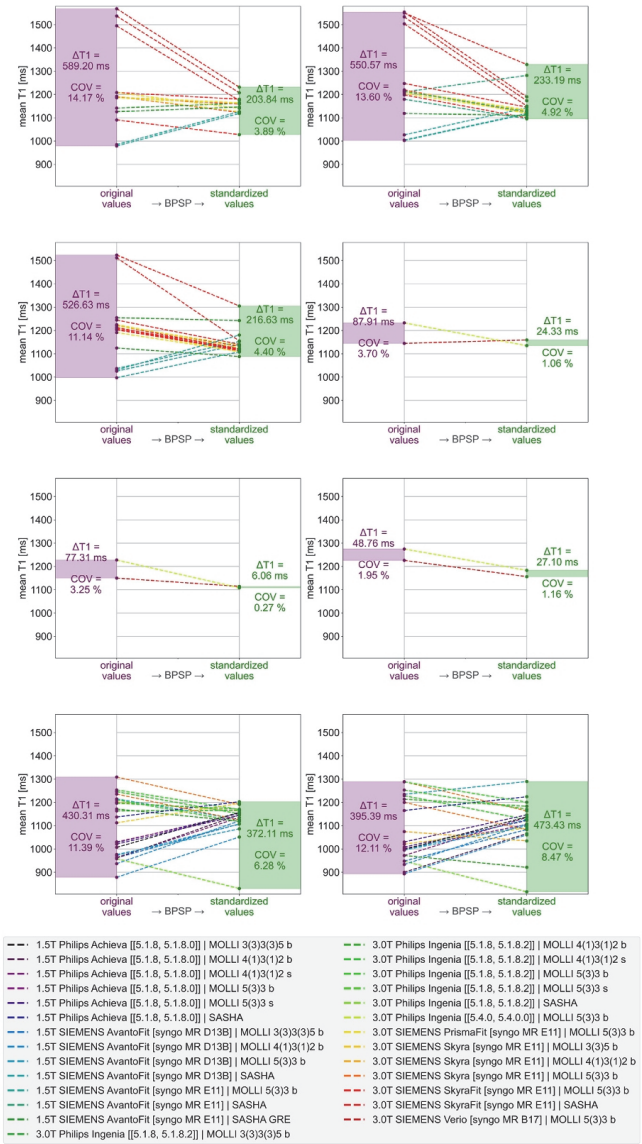

Fig. 6: Intra-subject progression plot – Eight healthy volunteers underwent at least two different acquisitions. The respective standardisation with the best performing standardisation pipeline (BPSP) is shown for each scanner and sequence combination. The violet bar on the left of each plot shows the value spread before and the green bar on the right the value spread after standardisation as well as the coefficient of variation (COV).

Discussion

In this work, we were able to show that a post-hoc standardisation of parametric T1 maps in CMR is feasible while maintaining disease differentiability according to sensitivity and specificity as in intra-scanner-intra-sequence scenarios. We could further show that the choice of the standardisation pipeline settings was crucial for the overall performance. The considered CPs of age, sex, scanner and sequence revealed the LSVR regression on relative values in the cascaded mode with two bins and the agglomerative single clustering as the BPSP among 240 tested standardisation pipelines with respect to the HTE of native, mid-ventricular parametric T1 maps in CMR. The implemented MARISSA including a GUI is made available with this work for further development and evaluation. In the following, a detailed discussion about the Dataset, the results of Confounding Parameters Impact Estimation (CPIE), Best Performing Standardisation Pipeline (BPSP), Diagnostic Implication (DI) and Implementation is presented and ends up with the Limitations and a short Conclusion.

Dataset

The included datasets were retrospectively collected from available data with a diagnosis of either Healthy, HCM or AMY. The data were not filtered except for native, artifact-free, midventricular slices. Consequently, the data were not balanced according to scanner-sequence combinations. The examination of both patient cohorts was mainly performed on 1.5T Siemens scanners with less sequence variations resulting in, contrary to literature,[15] lower average T1 values than HTE before standardisation. Other established manufacturers were not included due to missing access within the Berlin CMR research network.[10] The considered patient cohorts HCM and AMY reflect only two CVDs with expected significant higher T1 values. The diagnostic performance on other, more subtle, CVDs requires future investigation.

Regarding the segmentation, different strategies exist.[11,12] While the segmentation of the septum is more precise,[4] it lacks the majority of the myocardial voxels. Therefore, this work used a full circular segmentation of the myocardium at the cost of an increasing standard deviation in the T1 values compared to septal segmentation only.[12]

Confounding Parameters Impact Estimation (CPIE)

In this work, the CPIE trained regression models to estimate the bias between apparent and target T1 values based on a healthy volunteer cohort. Subjects with CVDs were excluded in this step, as those would have an unintentional influence towards either higher or lower T1 values.[25] The CPIE, as central part of the post-hoc standardisation pipeline, enables the comparability of parametric T1 mapping. Although multiple studies were published on reference values for parametric T1 mapping over the last decade, those were only valid in a specific cohort and technical setting.[12,26,27]

The z-Score transformation into a unitless value domain enabled as the solely established approach the comparability across different CP environments.[28] However, the z-Score calculation is based on a local healthy reference cohort. Consequently, a healthy cohort examination is not only performed after initial operation but always necessary after a hard- or software change of the MR scanner that perturbs the T1 value distribution of the healthy cohort.[28] Although considered in the current guidelines,[4] this obstacle of additional effort and costs circumvent the establishment of the z-Score in the clinical routine. The inclusion of novel CP values or even CPs themselves requires, likewise the z-Score transformation, additional healthy volunteer examinations. However, thanks to the transfer learning capability of MARISSA, single site scans are sufficient to apply CP value standardisation on other sites. Consequently, the amount of healthy volunteer examinations and thereby the cost may be reduced with our standardisation approach compared to the z-Score. Nonetheless, an increasing amount of training data is necessary over time, which is limited by the accessibility and potential restrictions due to institutional or governmental law.

The number of necessary training data highly depends on the standardisation pipeline setting, especially the mode. The individual mode already revealed in this work, that two sequences could not be captured during training as all other CPs were expected to be constant. Hence, the individual mode is prone to the training data and most likely misses certain CP values. However, it allows for the best isolation of CP's influence. Compared to that, the cascaded mode depends on the order of the CPs as the first one works the same way as the individual mode while the last one can consider the whole training dataset as all other CPs are already standardised. As a consequence, inter-parameter correlations are partly considered, but some CP values might be missed as well if the training dataset or the order of the considered CPs is not well chosen. Finally, the ensemble mode considers all in one and catches all CP values that were given in the training data. The ensemble mode accounts for inter-parameter correlation best, but fails completely if a test dataset includes a categorical value that was not in the training dataset. The individual and cascaded mode on the contrary still standardise for all other CPs and can skip those that are unknown. As a rule of thumb, the number of necessary training data increases from ensemble to cascaded to individual mode in order to capture all CP values.

As this study is a proof-of-concept for a standardisation approach of parametric T1 maps in CMR, only a limited number of parameters were included. There are other parameters such as heart rate (HR),[9,29] the body-mass-index (BMI)[30] or the voxel size[31] that have a known relevant impact. However, BMI and HR

Articles

could not be included as the necessary information were not available in all datasets of this study due to data anonymisation and its retrospective character. Further, all three CPs are numerical variables, resulting in a much higher necessary amount of training data for the individual and cascaded mode. On the contrary, each numerical variable can be turned into a categorical one by clustering, for example the age can be divided into decades or the BMI into the groups proposed by the world health organization.[32]

The patient specific parameters age and sex were included as potential CPs, but showed an ambivalent impact in the literature. Roy et al.[11] revealed a significant whereas Dabir et al.[12] showed only a slight but not significant impact due to age and sex. Our results, in turn, emphasised a non-zero induced bias, however, much less than the scanner or sequence originated biases. If age and sex turn into neglectable CPs compared to other potential CPs, those may be excluded easily from the standardisation pipeline within MARISSA in the future.

The considered CPs in this work treated subject specific and technological variations only. However, post-processing procedure variations represent CPs in parametric T1 mapping as well.[8] The post processing comprises everything from image reconstruction[8] to segmentation procedures.[33] While reconstruction algorithms are not available from DICOM tags, the segmentation method of the myocardium is still based on expert agreement or in accordance to guidelines.[34] Although fully automated segmentation procedures exist,[16] manual adaptions are still necessary. Consequently, the segmentation method is currently a local specific, hard to definable CP. In this work, the segmentation was performed according to a pre-defined standard operating procedure. However, in future perspective, when including more training data, this must be considered, especially since artificial intelligence assisted segmentation procedures become ever more popular.[16,35]

Best performing standardisation pipeline (BPSP)

As the post-hoc standardisation pipeline can be computed in different settings, the results showed that the setting choice is crucial for the performance of the pipeline with respect to the COV as quality index. Some pipeline settings even showed a worsening of the COV compared to the unstandardised values assuming a non-suitable standardisation pipeline setting. Although the BPSP among 240 evaluated standardisation pipeline settings included a clustering into two bins, the performance gain of 0.17% due to the clustering was rather small compared to a COV reduction of up to 6.49% by a standardisation pipeline without clustering. In most cases the clustering even worsened the outcome. Therefore, the determination of the suitable regression-type, y-type and mode is most important whereas clustering into bins reflects a rather potential fine-tuning step.

Diagnostic Implication (DI)

The BPSP allowed for a statistically significant differentiation of the three cohorts: HTE, HCM and AMY. However, the sum of sensitivity and specificity across HTE and HCM was below 150% due to a high overlap and thus not sufficient for evidence according to literature.[22] This aligns with literature values on 3T scanners for Healthy and HCM that show also a significant difference but high overlap as in Liang et al. (1228.4 ± 42.7 ms vs. 1290.0 ± 64.3 ms),[36] Qin et al. (1240.0 ± 29.8 ms vs. 1308.0 ± 55.5 ms)[37] and Lavall et al. (1225 ± 21 ms vs. 1266 ± 44 ms).[15] However, as HCM has manifold morphologies due to a large variety of genotypes and risk factors, the disease state changes over time, which in turn affects the amount of diseased myocardial tissue.[38,39]

Baggiano et al.[40] showed a sensitivity of 85% and specificity of 87% when comparing Healthy with AMY that could be outperformed with our BPSP with a sensitivity of 95.83% and specificity of 91.67%. However, they were able to include 436 patients with amyloidosis, which naturally assumes a higher value spread in that patient cohort compared to our 24 included ones. The z-Score approach reached in the study by Kranzusch et al. an equivalent sensitivity of 96% but an improved specificity of 100%.[28]

In the discrimination of both patient groups, HCM and AMY, the sensitivity (87.5%) and specificity (98.35%) were in range of published literature values by Lavall et al. (100% / 97%),[15] Nam et al. (76.1% / 83.3%)[41] and Martinez-Naharro (86.54% / 80.36%).[42] Consequently, amyloidosis is reliably detectable after standardisation while HCM only in an advanced state.

When going from the global cohort perspective into the intra-subject view, the BPSP managed to decrease the COV within the same subject across different acquisitions. Although most acquisitions could be harmonised towards equal values, individual outliers remained after standardisation. Those outliers originated mainly either from a SASHA based sequence, which are assumed to have a higher accuracy but lower precision than MOLLI based sequences,[9] or already had unusual values for the specific field strength and sequence scheme setup. This, however, shows the limits of MARISSA pipelines. On the one hand, outlying or unusual values will remain outlying or unusual after standardisation and, on the other hand, imprecision cannot be improved. Consequently, the used data for training and testing need a high degree of precision. This does not only affect the used sequence variant but also demands highly controlled production process of the magnetic resonance imaging scanner. High tolerances undermine the generalisability of the proposed post-hoc standardisation pipeline approach. This susceptibility to imprecision is also shared by the z-Score approach whose usability is undermined by high fluctuations in the standard deviation.[43] The standard

deviation of measured T1 values can be minimised by solely segmenting the septal region[12] instead of the full midventricular myocardium as performed in this work. However, this segmentation strategy would miss the majority of the myocardial tissue.

Finally, it is important to mention, that the standardisation approach calculates the impact of a CP compared to a reference CP value. Consequently, the standardised T1 maps become comparable, but do not necessarily represent the true T1 relaxation value of the myocardial tissue. The reference sequence MOLLI 5(3)3 b for example is known to underestimate the true T1 value.[9]

Implementation

As MARISSA was fully implemented in Python, it can be installed and run on all major operating systems. Export functionality enables the sharing of trained standardisation pipelines either with or without data. Although tested for parametric T1 maps of the heart only, MARISSA is intended for usability among other quantitative methods, like parametric T2 maps,[44] or tissues, such as the liver.[45]

Furthermore, MARISSA is extensible in the future. Additional conceivable CPs can be entered into MARISSA via the GUI while novel clustering algorithms and regression models are easily implementable due to a standardised structure.

A subsequent development of MARISSA may include further adjustment options within the GUI. This comprises for example the hyperparameter setup of regression models and clustering algorithms beyond the standard setting or individual settings for each CP within a standardisation pipeline.

Limitations

The major limitation of this work is the unbalanced underlying dataset due to its retrospective design. The lack of further scanner-sequence combinations limits the generalisability of the proposed standardisation pipeline. The transfer learning capability of scanner-sequence combinations that are not reflected in the training data but captured by the BPSP requires further investigation. This work only contained two cases in the AMY cohort whose scanner-sequence combination were not reflected in the training data. Additionally, a more convincing DI requires more scanner-sequence variability in the considered patient groups. As this work is a proof-of-concept and includes anonymised data, further known relevant CPs were not included and should be considered in a future state. The inclusion of mid-ventricular slices only does not meet all manifold phenotypes of an HCM which may affect only various local regions rather than the whole ventricle.[46]

Conclusion

All in all, we were able to introduce the MARISSA to enable post-hoc standardisation pipelines for parametric T1 mapping in CMR. The diagnostic power after standardisation with the BPSP in our proof-of-concept were equivalent to those found in literature. The performance of the standardisation pipeline highly depends on the pipeline setting and the precision of the provided data. The current results give hope to improve comparability when adding more training data and considered CPs in the future.

Contributors

Darian Viezzer – Conceptualisation, Methodology, Data Curation, Formal Analysis, Project Administration, Software, Validation, Visualisation, Writing (original draft).
Thomas Hadler – Methodology, Formal Analysis, Software, Validation, Writing (review & editing).
Jan Gröschel – Data Curation, Formal Analysis, Writing (review & editing).
Clemens Ammann – Methodology, Software, Writing (review & editing).
Edyta Blaszczyk – Data Curation, Formal Analysis, Writing (review & editing).
Christoph Kolbitsch – Methodology, Writing (review & editing).
Simone Hufnagel – Data Curation, Writing (review & editing).
Riccardo Kranzusch-Groß – Data Curation, Writing (review & editing).
Steffen Lange – Conceptualisation, Methodology, Formal Analysis, Supervision, Writing (review & editing).
Jeanette Schulz-Menger – Conceptualisation, Data Curation, Formal Analysis, Supervision, Writing (review & editing).
All authors read and approved the final version of the manuscript and ensured it is the case. Darian Viezzer, Jan Gröschel and Edyta Blaszczyk accessed and verified the complete underlying data.

Data sharing statement

The trained standardisation pipelines are available on reasonable request by contacting the first (DV: darian-steven.viezzer@charite.de) or last (JSM: jeanette.schulz-menger@charite.de) author. Due to institutional law, datasets cannot be shared. The source code can be accessed via GitHub under the URL: https://github.com/DSV-CUB/MARISSA or in the Supplemental Material S2.[23]

Declaration of interests

The authors declare no competing interests and no usage of AI or AI assisted technologies for the scientific writing.

Acknowledgements

We thank Daniel Messroghli for contributing some of the datasets. Furthermore, we thank Theodora Chitiboi and Jens Wetzl from Siemens Healthcare GmbH for giving access to their deep learning model for the segmentation of parametric T1 maps. This study was supported by the BMBF (Bundesministerium für Bildung und Forschung)/DZHK (German Centre for Cardiovascular Research) via project FKZ81Z0100208.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2024.105055.

References

1 Arbelo E, Protonotarios A, Gimeno JR, et al. 2023 ESC Guidelines for the management of cardiomyopathies: developed by the task force on the management of cardiomyopathies of the European Society of Cardiology (ESC). *Eur Heart J*. 2023;44(37):3503–3626.
2 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.
3 Guo R, Weingärtner S, Šiurytė P, et al. Emerging techniques in cardiac magnetic resonance imaging. *J Magn Reson Imag*. 2022;55(4):1043–1059.

Articles

4 Messroghli DR, Moon JC, Ferreira VM, et al. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: a consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *J Cardiovasc Magn Reson*. 2017;19(1):75.
5 Kammerlander AA, Marzluf BA, Zotter-Tufaro C, et al. T1 mapping by CMR imaging: from histological validation to clinical implication. *JACC Cardiovasc Imaging*. 2016;9(1):14–23.
6 Warnica W, Al-Arnawoot A, Stanimirovic A, et al. Clinical impact of cardiac MRI T1 and T2 parametric mapping in patients with suspected cardiomyopathy. *Radiology*. 2022;305(2):319–326.
7 Merlo M, Gagno G, Baritussio A, et al. Clinical application of CMR in cardiomyopathies: evolving concepts and techniques. *Heart Fail Rev*. 2023;28(1):77–95.
8 Ogier AC, Bustin A, Cochet H, Schwitter J, van Heeswijk RB. The road toward reproducibility of parametric mapping of the heart: a technical review. *Front Cardiovasc Med*. 2022;9 [cited 2023 Jul 31] Available from: https://www.frontiersin.org/articles/10.3389/fcvm.2022.876475.
9 Kellman P, Hansen MS. T1-mapping in the heart: accuracy and precision. *J Cardiovasc Magn Reson*. 2014;16(1):2.
10 Gröschel J, Trauzeddel RF, Müller M, et al. Multi-site comparison of parametric T1 and T2 mapping: healthy travelling volunteers in the Berlin research network for cardiovascular magnetic resonance (BER-CMR). *J Cardiovasc Magn Reson*. 2023;25(1):47.
11 Roy C, Slimani A, de Meester C, et al. Age and sex corrected normal reference values of T1, T2 T2* and ECV in healthy subjects at 3T CMR. *J Cardiovasc Magn Reson*. 2017;19(1):72.
12 Dabir D, Child N, Kalra A, et al. Reference values for healthy human myocardium using a T1 mapping methodology: results from the International T1 Multicenter cardiovascular magnetic resonance study. *J Cardiovasc Magn Reson*. 2014;16(1):69.
13 Roujol S, Weingärtner S, Foppa M, et al. Accuracy, precision, and reproducibility of four T1 mapping sequences: a head-to-head comparison of MOLLI, ShMOLLI, SASHA, and SAPPHIRE. *Radiology*. 2014;272(3):683–689.
14 Child N, Suna G, Dabir D, et al. Comparison of MOLLI, shMOLLLI, and SASHA in discrimination between health and disease and relationship with histologically derived collagen volume fraction. *Eur Heart J Cardiovasc Imaging*. 2018;19(7):768–776.
15 Lavall D, Vosshage NH, Geßner R, et al. Native T1 mapping for the diagnosis of cardiac amyloidosis in patients with left ventricular hypertrophy. *Clin Res Cardiol*. 2023;112(3):334–342.
16 Viezzer D, Hadler T, Ammann C, et al. Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance. *Sci Rep*. 2023;13(1):2103.
17 Smith G. 8–simple regression. In: Smith G, ed. *Essential statistics, regression, and econometrics*. 2nd ed. Boston: Academic Press; 2015:219–259 [cited 2023 Nov 1] Available from: https://www.sciencedirect.com/science/article/pii/B978012803459000008X.
18 Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: a library for large linear classification. *J Mach Learn Res*. 2008;9(61):1871–1874.
19 Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
20 Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63(1):3–42.
21 Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12(85):2825–2830.
22 Power M, Fell G, Wright M. Principles for high-quality, high-value testing. *BMJ Evid Based Med*. 2013;18(1):5–10.
23 Viezzer DS, Hadler T, Ammann C. *MARISSA–magnetic resonance imaging software for standardization*. Zenodo; 2024 [cited 2024 Feb 9]. Available from: https://zenodo.org/records/10640021.
24 NEMA PS3/ISO 12052, digital imaging and Communications in medicine (DICOM) standard, National Electrical Manufacturers Association, Rosslyn, VA, USA. Available from: http://www.dicomstandard.org/.
25 Haaf P, Garg P, Messroghli DR, Broadbent DA, Greenwood JP, Plein S. Cardiac T1 Mapping and Extracellular Volume (ECV) in clinical practice: a comprehensive review. *J Cardiovasc Magn Reson*. 2016;18(1):89.
26 Shaw M, Ojha V, Ganga KP, et al. Reference values of myocardial native T1 and T2 mapping values in normal Indian population at 1.5 Tesla scanner. *Int J Cardiovasc Imag*. 2022;38(11):2403–2411.
27 Abdullah H, Abdul Kadir A, Sha'rani MA, Tan SH, Othman N. Pilot study to establish the myocardium native T1 mapping reference values in 1.5 Tesla cardiac magnetic resonance of healthy Malaysians. *Egypt J Radiol Nucl Med*. 2023;54(1):98.
28 Kranzusch R, aus dem Siepen F, Wiesemann S, et al. Z-score mapping for standardized analysis and reporting of cardiovascular magnetic resonance modified Look-Locker inversion recovery (MOLLI) T1 data: normal behavior and validation in patients with amyloidosis. *J Cardiovasc Magn Reson*. 2020;22(1):6.
29 Granitz M, Motloch LJ, Granitz C, et al. Comparison of native myocardial T1 and T2 mapping at 1.5T and 3T in healthy volunteers : reference values and clinical implications. *Wien Klin Wochenschr*. 2019;131(7–8):143–155.
30 Zhao H, Huang R, Jiang M, et al. Myocardial tissue-level characteristics of adults with metabolically healthy obesity. *JACC Cardiovasc Imaging*. 2023;16(7):889–901.
31 Hufnagel S, Metzner S, Kerkering KM, et al. 3D model-based super-resolution motion-corrected cardiac T1 mapping. *Phys Med Biol*. 2022;67(24):245008.
32 *WHO Expert Committee on Physical Status: the Use and Interpretation of Anthropometry. 1993: Geneva S, Organization WH. Physical status : the use of and interpretation of anthropometry, report of a WHO expert committee*. World Health Organization; 1995 [cited 2023 Aug 26]. Available from: https://apps.who.int/iris/handle/10665/37003.
33 Zange L, Muehlberg F, Blaszczyk E, et al. Quantification in cardiovascular magnetic resonance: agreement of software from three different vendors on assessment of left ventricular function, 2D flow and parametric mapping. *J Cardiovasc Magn Reson*. 2019;21(1):12.
34 Hadler T, Wetzl J, Lange S, et al. Introduction of Lazy Luna an automatic software-driven multilevel comparison of ventricular function quantification in cardiovascular magnetic resonance imaging. *Sci Rep*. 2022;12(1):6629.
35 Kalapos A, Szabó L, Dohy Z, et al. Automated T1 and T2 mapping segmentation on cardiovascular magnetic resonance imaging using deep learning. *Front Cardiovasc Med*. 2023;10:1147581.
36 Liang L, Wang X, Yu Y, et al. T1 mapping and extracellular volume in cardiomyopathy showing left ventricular hypertrophy: differentiation between hypertrophic cardiomyopathy and hypertensive heart disease. *Int J Gen Med*. 2022;15:4163–4173.
37 Qin L, Min J, Chen C, et al. Incremental values of T1 mapping in the prediction of sudden cardiac death risk in hypertrophic cardiomyopathy: a comparison with two guidelines. *Front Cardiovasc Med*. 2021;8 [cited 2023 Aug 26] Available from: https://www.frontiersin.org/articles/10.3389/fcvm.2021.661673.
38 Neubauer S, Kolm P, Ho CY, et al. Distinct subgroups in hypertrophic cardiomyopathy in the NHLBI HCM registry. *J Am Coll Cardiol*. 2019;74(19):2333–2345.
39 Harper AR, Goel A, Grace C, et al. Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. *Nat Genet*. 2021;53(2):135–142.
40 Baggiano A, Boldrini M, Martinez-Naharro A, et al. Noncontrast magnetic resonance for the diagnosis of cardiac amyloidosis. *JACC Cardiovasc Imaging*. 2020;13(1, Part 1):69–80.
41 Nam BD, Kim SM, Jung HN, Kim Y, Choe YH. Comparison of quantitative imaging parameters using cardiovascular magnetic resonance between cardiac amyloidosis and hypertrophic cardiomyopathy: inversion time scout versus T1 mapping. *Int J Cardiovasc Imag*. 2018;34(11):1769–1777.
42 Martinez-Naharro A, Kotecha T, Norrington K, et al. Native T1 and extracellular volume in transthyretin amyloidosis. *JACC Cardiovasc Imaging*. 2019;12(5):810–819.
43 Popescu IA, Werys K, Zhang Q, et al. Standardization of T1-mapping in cardiovascular magnetic resonance using clustered structuring for benchmarking normal ranges. *Int J Cardiol*. 2021;326:220–225.
44 Wiesmueller M, Wuest W, Heiss R, Treutlein C, Uder M, May MS. Cardiac T2 mapping: robustness and homogeneity of standardized in-line analysis. *J Cardiovasc Magn Reson*. 2020;22(1):39.
45 Obmann VC, Berzigotti A, Catucci D, et al. T1 mapping of the liver and the spleen in patients with liver fibrosis—does normalization to the blood pool increase the predictive value? *Eur Radiol*. 2021;31(6):4308–4318.
46 Soler R, Méndez C, Rodríguez E, Barriales R, Ochoa JP, Monserrat L. Phenotypes of hypertrophic cardiomyopathy. An illustrative review of MRI findings. *Insights Imaging*. 2018;9(6):1007–1020.

Curriculum Vitae

My Curriculum Vitae does not appear in the electronic version of my doctoral thesis for data protection reasons.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Doktorarbeit nicht veröffentlicht.

My Curriculum Vitae does not appear in the electronic version of my doctoral thesis for data protection reasons.

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Doktorarbeit nicht veröffentlicht.

Publication list

First Authorship

Viezzer, D., Hadler, T., Gröschel, J., Ammann, C., Blaszczyk, E., Kolbitsch, C., Hufnagel, S., Kranzusch-Groß, R., Lange, S. & Schulz-Menger, J., *Post-hoc standardisation of parametric T1 maps in cardiovascular magnetic resonance imaging: a proof-of-concept*, eBioMedicine, 2024, IF: 11.1

Viezzer, D., Hadler, T., Ammann, C., Blaszczyk, E., Fenski, M., Grandy, T. H., Wetzl, J., Lange, S. & Schulz-Menger, J., *Introduction of a cascaded segmentation pipeline for parametric T1 mapping in cardiovascular magnetic resonance to improve segmentation performance*, Scientific Reports, 2023, IF: 5.0

Co-Authorship

Gröschel, J., Kuhnt, J., **Viezzer, D.**, Hadler, T., Hormes, S., Barckow, P., Schulz-Menger, J. & Blaszczyk, E., *Comparison of manual and artificial intelligence based quantification of myocardial strain by feature tracking-a cardiovascular MR study in health and disease*, European Radiology, 2024, IF: 5.9

Gröschel, J., Trauzeddel, R.-F., Müller, M., von Knobelsdorff-Brenkenhoff, F., **Viezzer, D.**, Hadler, T., Blaszczyk, E., Daud, E. & Schulz-Menger, J., *Multi-site comparison of parametric T1 and T2 mapping: healthy travelling volunteers in the Berlin research network for cardiovascular magnetic resonance (BER-CMR)*, Journal of Cardiovascular Magnetic Resonance, 2023, IF: 6.4

Hadler, T., Ammann, C., Wetzl, J., **Viezzer, D.**, Gröschel, J., Fenski, M., Abazi, E., Lange, S., Hennemuth, A. & Schulz-Menger, J., *Lazy Luna: Extendible software for multilevel reader comparison in cardiovascular magnetic resonance imaging*, Computer Methods and Programs in Biomedicine, 2023, IF: 6.1

Riazy, L., Däuber, S., Lange, S., **Viezzer, D.**, Ott, S., Wisemann, S., Blaszczyk, E., Mühlberg, F., Zange, L. & Schulz-Menger, J., *Translating principles of quality control to cardiovascular magnetic resonance: assessing quantitative parameters of the left ventricle in a large cohort*, Scientific Reports, 2023, IF: 5.0

Gröschel, J., Bhoyroo, Y., Blaszczyk, E., Trauzeddel, R.-F., **Viezzer, D.**, Saad, H., Fenski, M. & Schulz-Menger, J., *Different Impacts on the Heart After COVID-19 Infection and Vaccination: Insights From Cardiovascular Magnetic Resonance*, Frontiers in Cardiovascular Medicine, 2022, IF: 3.6

Gröschel, J., Ammann, C., Zange, L., **Viezzer, D.**, Forman, C., Schmidt, M., Blaszczyk, E. & Schulz-Menger, J., *Fast acquisition of left and right ventricular function parameters applying cardiovascular magnetic resonance in clinical routine – validation of a 2-shot compressed sensing cine sequence*, Scandinavian Cardiovascular Journal, 2022, IF: 2.0

Fenski, M., Grandy, T. H., **Viezzer, D.**, Kertusha, S., Schmidt, M., Forman, C. & Schulz-Menger J., *Isotropic 3D compressed sensing (CS) based sequence is comparable to 2D-LGE in left ventricular scar quantification in different disease entities*, International Journal of Cardiovascular Imaging, 2022, IF: 2.1

Devriendt, T., Ammann, C., Asselbergs, F. W., Bernier, A., Costas, R., Friedrich, M., Gelpi, J., Jarvelin, M., Kuulasmaa, K., Lekadir, K., Mayrhofer, M., Papez, V., Pasterkamp, G., Petersen, S., Schmidt, C., Schulz-Menger, J., Söderberg, S., Shabani, M., Veronesi, G., **Viezzer, D.** & Borry, P., *An agenda-setting paper on data sharing platforms: euCanSHare workshop*, Open Research Europe, 2021, IF: none

Conference Contributions

Viezzer, D., Hadler, T., Gröschel, J., Ammann, C., Blaszczyk, E., Kolbitsch, C., Lange, S. & Schulz-Menger J., *Post-hoc Standardization of Parametric T1 Mapping in Cardiovascular Magnetic Resonance Imaging: a first proof-of-concept*, CMR, 2024

Viezzer, D., Hadler, T., Blaszczyk, E., Fenski, M., Gröschel, J., Lange, S. & Schulz-Menger J., *Improving U-Net based segmentation in cardiac parametrical T1 mapping by incorporating bounding box information*, ISMRM, 2022

Viezzer, D., Lange, S. & Schulz-Menger, J., *Temperature Dependency of Parametric Mapping in Cardiovascular Magnetic Resonance Imaging validated in a 3 Tesla T1MES Phantom*, IEEE ISBI, 2021

Acknowledgments

Per aspera ad astra

The completion of a doctoral thesis is characterized by deep valleys and dizzying heights. Sometimes one loses one's way and then it is important to get back on track. In this point I want to specially thank my first supervisor Univ.-Prof. Dr. med Jeanette Schulz-Menger for her guidance through the project, her empathic understanding when I got stuck and her faith in me. I would never have dreamed of completing a doctoral thesis before. She has accompanied me since my Master studies and without her I would not stand where I am including all the ideas, discussion, criticisms, motivations and helps we have shared.

Numquam solus es

This work would never be possible without the contribution of many additional people. I would like to specially mention Prof. Dr. Steffen Lange, who took over the role as a second supervisor. He provided great support in technical questions and structuring thoughts. His abstract way of thinking was an awesome source of inspiration. As I cannot list all of them, I would like to further thank all my co-authors and acknowledged individuals in the published work. Their contributions were essential to the realization of this work. Last but not least, I am grateful for the whole team of the working group on cardiovascular magnetic resonance imaging at the Charité – Universitätsmedizin Berlin and the associated team at the Helios Klinikum Berlin-Buch. Everyone made, either directly or indirectly, an impact on me and this work.

Aurea mediocritas

Even if it fades into nothing, I would like to thank all the musicians I listened to through long lasting programming sessions, helping me to stay motivated and concentrated. Finally, I would like to pay tribute to the most important people in my life. As a balance, companion is inevitable and through the journey of this doctoral thesis, I would like to thank dearly my family, especially my parents Jacqueline Meier and Steven Viezzer, my grandparents Eileen Everton and Renato Viezzer, my closely relatives Andreas Thrun, Gabriele Töpper and Isabell Töpper, my partner Dana Diack, my cousin Matilda Iosco, my beloved friends, particularly my longest known companions Daniel Kammholz and Anna Lamprecht, and my fluffy cats Cosma, Luna and Lupin.