Quality control-driven framework for reliable automated segmentation of cardiac magnetic resonance LGE and VNE images

Ricardo A. Gonzales[1], Daniel H. Ibáñez[1,2], Evan Hann[1], Iulia A. Popescu[1], Matthew K. Burrage[1,3], Yung P. Lee[1], İbrahim Altun[1], William S. Weintraub[4], Raymond Y. Kwong[5], Christopher M. Kramer[6], Stefan Neubauer[1], Vanessa M. Ferreira[1], Qiang Zhang[1], and Stefan K. Piechnik[1]

[1] Radcliffe Department of Medicine, University of Oxford, Oxford, UK
[2] Artificio, MA, USA
[3] Faculty of Medicine, University of Queensland, Brisbane, Australia
[4] MedStar Health Research Institute, Georgetown University, DC, USA
[5] Cardiovascular Division, Harvard Medical School, MA, USA
[6] Department of Medicine, University of Virginia Health System, VA, USA

1 Introduction

Accountable automated cardiovascular imaging analysis is pivotal for accurate diagnosis and treatment decisions, yet remains a challenge due to limited training data and lack of quality control. One such application is in late gadolinium enhancement (LGE) cardiovascular magnetic resonance (CMR) imaging, the gold standard for non-invasive myocardial tissue characterisation [1]. To circumvent the manual and often subjective process of segmenting left ventricular (LV) myocardium, automated segmentation methods have emerged [2]. However, the clinical translation has been limited by scarcity of high-quality training data [3], and the potential for unflagged segmentation errors [4].

Transfer learning, domain adaptation, and data augmentation techniques have been proposed to mitigate data scarcity [5]. Particularly, generative adversarial networks (GANs) have been employed to augment data through the generation of synthetic variants [6]. Yet, their usage in medical applications necessitates rigorous clinical validation [7]. Moreover, efficient quality control mechanisms are required to detect and flag segmentation errors [8]. While recent efforts for automatic error-flagging have shown promise [9], a quality control pipeline specific to LGE segmentation is still lacking. We address these challenges by presenting an automated quality control to enhance the clinical translation of LGE segmentation, directly applicable to GAN-generated virtual native enhancement (VNE) images [10].

2 Methods

The study used a development dataset of 4,716 LGE images from 1,363 patients, sourced from the Hypertrophic Cardiomyopathy Registry study [11], the University of Oxford Centre for Clinical Magnetic Resonance Research, and the Oxford

Acute Myocardial Infarction study [12]. The dataset was expanded using a conditional generative adversarial network [13] to create VNE images [10], which simulated LGE images without the need for gadolinium.

A quality control-driven (QCD) ensemble framework was developed [9], leveraging multiple U-Nets [14] for more accurate and robust LGE and VNE segmentation. It used statistical rank filters to create a diverse pool of candidate segmentations and a quality scoring mechanism to predict the Dice Similarity Coefficient (DSC) for each segmentation. The final segmentation was selected by identifying the candidate with the highest predicted DSC.

The dataset was augmented with VNE technology, using co-located cines and ShMOLLI T1 maps, resulting in 3,541 VNE images. The augmented dataset was partitioned into training (85%), validation (7.5%), and test (7.5%) sets. The segmentation models were optimised with the Adam method, and a linear regressor was fit for each candidate segmentation. The performance of the ensemble framework was evaluated on both LGE and VNE test datasets, measuring segmentation accuracy by DSC and prediction accuracy by mean absolute error and binary classification accuracy, with a DSC threshold of 0.7.

3 Results

The results show that the QCD framework was able to successfully and efficiently segment the LV myocardium on LGE and VNE images. The framework demonstrated robust performance on both types of images with a mean DSC of 0.845 ± 0.075 for LGE and 0.845 ± 0.071 for VNE. The mean absolute error for the predicted DSC was low at 0.043, and the binary classification accuracy was high at 0.951, confirming the effectiveness of the ensemble framework. Without the augmented data and the QCD approach, using an individual U-Net, the segmentation performance dropped to 0.836 ± 0.082 and 0.838 ± 0.075, for LGE and VNE, respectively, and without the error-flagging capability. Figure 1 exhibits representative test cases of the QCD framework on LGE images for true positive, true negative, false positive and false negative cases.

4 Discussion

This study demonstrates the effectiveness of an automated quality control mechanism for improving the accuracy and reliability of LGE segmentation, directly applicable to GAN-generated VNE data. It emphasises the importance of VNE data in training processes and highlights the framework's robustness with LGE and VNE data. This work represents a key step in enabling faster, reliable diagnoses of myocardial damage.

Data scarcity is a major issue in developing reliable deep learning models for medical image segmentation, particularly LGE segmentation. To overcome this, the study uses VNE images to augment existing LGE data, resulting in a more robust model that can better handle diverse and complex clinical cases. The study focuses on integrating an automated quality assurance framework using

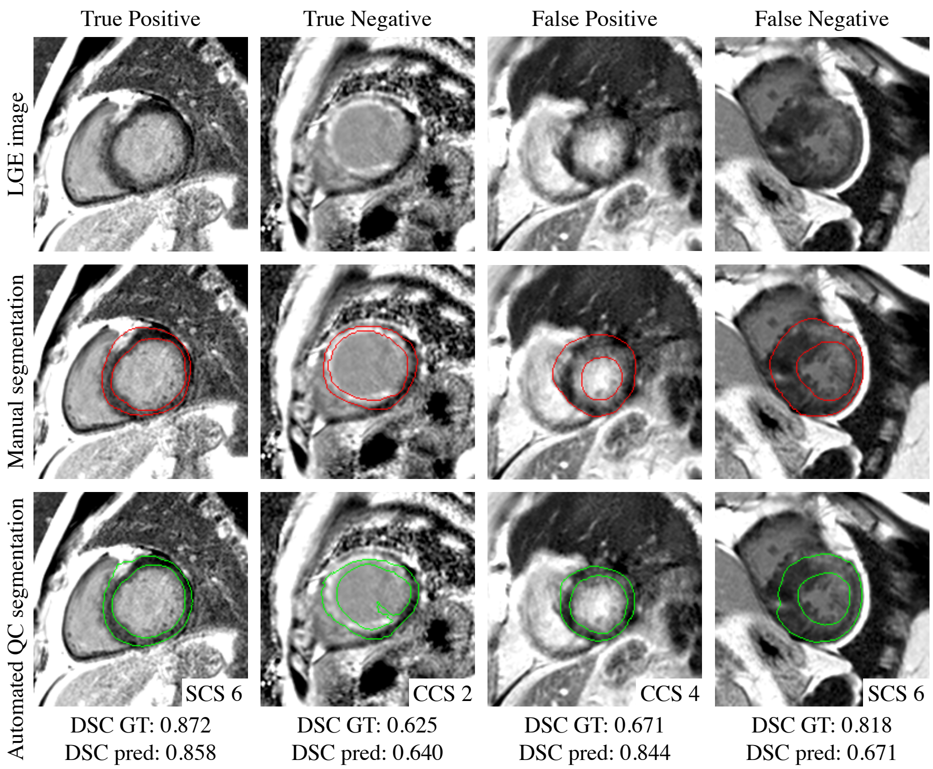

Fig. 1. Examples of true positive (93.9%), true negative (1.9%), false positive (2.3%) and false negative (1.9%) for predicted quality-controlled (QC) segmentations in late gadolinium enhanced (LGE) images. The left ventricular myocardium is manually segmented in red and automatically segmented in green, from different single (SCS) and combined candidate segmentation (CCS) models. The corresponding observed ground-truth (GT) Dice Similarity Coefficient (DSC) and predicted DSC are provided at the bottom. Note: This figure is also included in an extended manuscript in Frontiers in Cardiovascular Medicine [15].

a traditional encoder-decoder U-Net architecture. The QCD strategy offered reliable quality predictions, critical for clinical decisions, proving to be more effective than the Monte Carlo-based quality assurance scheme [16]. Future work will consider newer network architectures, advanced pre-processing schemes, and evaluating scar burden.

The QCD ensemble framework's clinical implications are substantial as it introduces an automated quality control mechanism to LGE segmentation for the first time, increasing accuracy and reliability. The QCD strategy allows identification and refinement of suboptimal segmentations, streamlining the diagnostic process, reducing variability, boosting clinician confidence, and improving patient outcomes. It also sets the stage for increased clinical adoption.

5 Conclusion

We introduce a new methodology for automated LGE segmentation, addressing limited training data and the lack of clinical quality control. Using GAN-generated VNE images and an automated quality control system, we display the possibility for enhancing performance and dependability in automated segmentation. This framework can seamlessly integrate into clinical practice, presenting a robust tool for clinicians to address myocardial damage.

References

1. Kim, R.J., et al.: The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. N Engl J Med (2000)
2. Zhuang, X., et al.: Cardiac segmentation on late gadolinium enhancement MRI: A benchmark study from multi-sequence cardiac MR segmentation challenge. Med Image Anal (2022)
3. Willemink, M.J., et al.: Preparing medical imaging data for machine learning. Radiology (2020)
4. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis. IEEE Trans Med Imaging (2018)
5. Tajbakhsh, N., et al.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Med Image Anal (2020)
6. Al Khalil, Y., et al.: On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images. Med Image Anal (2023)
7. Chen, R.J., et al.: Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng (2021)
8. Galati, F., et al.: From accuracy to reliability and robustness in cardiac magnetic resonance image segmentation: A review. Appl Sci (2022)
9. Hann, E., et al.: Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. Med Image Anal (2021)
10. Zhang, Q., et al.: Artificial intelligence for contrast-free MRI: Scar assessment in myocardial infarction using deep learning–based virtual native enhancement. Circulation (2022)
11. Kramer, C.M., et al.: Hypertrophic Cardiomyopathy Registry: The rationale and design of an international, observational study of hypertrophic cardiomyopathy. Am Heart J (2015)
12. Scarsini, R., et al.: Coronary microvascular dysfunction assessed by pressure wire and CMR after STEMI predicts long-term outcomes. JACC: Cardiovasc Imaging (2021)
13. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: 2017 IEEE CVPR (2017)
14. Ronneberger, O., et al.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI 2015 (2015)
15. Gonzales, R.A., et al.: Quality control-driven deep ensemble for accountable automated segmentation of cardiac magnetic resonance LGE and VNE images. Front Cardiovasc Med (2023)
16. Hann, E., et al.: Ensemble of deep convolutional neural networks with Monte Carlo dropout sampling for automated image segmentation quality control and robust deep learning using small datasets. In: MIUA 2021