

Improving the quality of information systems for diagnosis: an approach to detecting and correcting segmentation errors

Bruno F. Raquel¹, Matheus A. O. Ribeiro¹, Fátima L. S. Nunes¹

¹Laboratório de Aplicações de Informática em Saúde,
Escola de Artes, Ciências e Humanidades, Universidade de São Paulo
Rua Arlindo Bértio, 1000, São Paulo – SP, Brasil, 03828-000

{brunofraquel, matheus.alberto.ribeiro, fatima.nunes}@usp.br

Abstract. Research Context: *The development of automated Information Systems (IS) capable of segmenting the left ventricle (LV) in cardiac magnetic resonance imaging (MRI) and estimating clinically relevant biomarkers is fundamental to support diagnostic decision-making. Scientific and/or Practical Problem:* *Many automated IS based on deep learning (DL) are not fully reliable and lack dedicated modules for error detection, which makes them dependent on constant manual inspection and correction. Proposed Solution and/or Analysis:* *We propose a post-processing method for IS that automatically detects and corrects segmentation errors in LV cardiac MRI produced by DL systems. Detection is performed by combining metrics computed between consecutive time frames of MRI to identify inconsistent segmentations in the temporal dimension; the correction step reconstructs the problematic frame by interpolating nearby segmentations. Related IS Theory:* *This work is grounded in the perspectives of Information Processing Theory. Research Method:* *The method was validated on LV segmentations containing both artificially generated and real DL errors, using the LVQuan19 dataset with reference segmentations for all time instants. Summary of Results:* *The method achieved detection performance with F1-score values up to 0.99 on real data, particularly for severe errors. Regarding the correction step, the selected strategy effectively improved segmentation consistency, achieving Dice coefficient values close to 0.95, indicating excellent agreement with reference segmentations. Contributions and Impact to IS area:* *This study contributes to the Information Systems field by introducing a method that improves the reliability of automated computer-aided systems for diagnosis. Although validated in the context of LV segmentation, the proposed approach can be applied to other domains where sequential data consistency is critical.*

1. Introdução

A crescente demanda e o aumento do número de imagens por exame tornam a interpretação manual lenta e sujeita a variabilidade, o que reforça a importância de sistemas de apoio à decisão clínica (CAD) capazes de automatizar etapas da análise de imagem. Em particular, a segmentação automática é um componente central desses sistemas, pois favorece a extração de biomarcadores que suportam a decisão clínica. Sistemas de informação (SI) e ferramentas de apoio à decisão têm se tornado cada vez mais presentes na rotina clínica, especialmente como recursos de auxílio ao diagnóstico [Guetari et al. 2023]. Diversos trabalhos têm sido publicados

nesse campo, com aplicações em múltiplas modalidades e condições, como retinopatia diabética [Asiri et al. 2019], tumor mamário [Moon et al. 2020] e câncer de tireoide [Anari et al. 2022]. Na cardiologia, a Ressonância Magnética Cardíaca (RMC) é considerada o padrão-ouro para avaliação estrutural e funcional do coração [Bernard et al. 2018], e o ventrículo esquerdo (VE) é alvo rotineiro de segmentação por fornecer biomarcadores clínicos essenciais. As segmentações realizadas por especialistas são tomadas como referência, sendo consideradas corretas e utilizadas como *ground truth* (GT) para validação de métodos automatizados.

Métodos baseados em aprendizado profundo (AP) vêm apresentando desempenho elevado em competições e bases públicas para a segmentação do VE [Shoaib et al. 2023], com resultados, em geral, muito próximos às segmentações realizadas por especialistas [Bernard et al. 2018]. Contudo, avanços em métricas globais amplamente utilizadas, como a métrica de sobreposição Dice, nem sempre se traduzem em segmentações clinicamente válidas. Comparações mostram que mesmo o melhor método de segmentação pode gerar erros em 82% dos exames — valor obtido após verificação manual das segmentações, não evidenciado pelas métricas utilizadas — o que pode comprometer a obtenção de biomarcadores corretos [Bernard et al. 2018]. A detecção e a correção automáticas dessas falhas são ainda dificultadas pela falta de métricas adequadas, pois a avaliação padrão depende de GTs que não se aplicam a novos exames da rotina clínica. Essas limitações ressaltam a necessidade de abordagens que não apenas apresentem métricas de desempenho elevadas, mas que também não gerem erros.

O diagnóstico de alterações funcionais do coração depende da análise do VE ao longo de múltiplos instantes do ciclo cardíaco, uma vez que medidas temporais de volume e morfologia são frequentemente utilizadas como biomarcadores. Erros em segmentações sequenciais degradam essas métricas, podendo assim afetar o diagnóstico de diversas patologias ao criar discrepâncias que se confundem com sinais clínicos reais. Um exemplo são as dissincronias ventriculares, condição em que diferentes regiões do VE se contraem de forma descoordenada, e cuja identificação depende da análise do ventrículo ao longo do ciclo cardíaco. Na ausência de um controle de qualidade automático eficaz, o especialista não dispõe de um mecanismo confiável para distinguir se uma inconsistência observada decorre de uma anomalia anatômica ou de uma falha de segmentação, o que tende a aumentar a taxa de falsos positivos e a necessidade de revisões manuais, consumindo recursos humanos. Por esses motivos, é desejável que métodos automáticos incorporem mecanismos de detecção e correção de erros sequenciais em segmentações do VE, de modo a preservar a confiabilidade dos biomarcadores extraídos.

Diante desse cenário, o presente trabalho propõe um método de pós-processamento para segmentações do VE em exames médicos, capaz de detectar e corrigir erros em sequências temporais do ciclo cardíaco, visando melhorar a qualidade das segmentações e favorecer a extração de biomarcadores mais confiáveis. As principais contribuições deste estudo são: (i) o desenvolvimento de um algoritmo automático para detecção de erros em sequências de segmentações do VE; e (ii) a implementação de uma estratégia de correção que preserva a consistência temporal entre instantes do ciclo cardíaco. O método de pós-processamento é proposto como parte de um SI para o diagnóstico de cardiomiopatias. Mais detalhes sobre o sistema CAD ao qual o método foi incorporado serão apresentados na Seção 3.1. Embora desenvolvido consi-

derando segmentações do VE, o método pode ser integrado como um módulo de pós-processamento à saída de qualquer método de segmentação (incluindo soluções baseadas em AP), tornando-o potencialmente aplicável a outras modalidades de imagem e contextos clínicos nos quais deseja-se avaliar a consistência sequencial de segmentações.

Inserido no domínio da saúde, o estudo aborda um problema com implicações organizacionais e clínicas para SI: a necessidade de aumentar a confiabilidade da informação produzida por sistemas de apoio ao diagnóstico baseados em segmentação automática em CAD. No contexto dos grandes desafios do GranDSI-BR, a abordagem proposta contribui para a confiabilidade de SI que possuam processamento intensivo de dados ao incorporar mecanismos automáticos de controle de qualidade que fortalecem a tomada de decisão clínica. Sob a perspectiva de "Processos", o método aumenta a robustez dos fluxos de trabalho ao mitigar falhas que demandariam revisão manual; no eixo de "Tecnologias", introduz um módulo de pós-processamento escalável e mais independente de inspeção manual; e, no eixo de "Pessoas", reduz a carga cognitiva do especialista e aumenta a confiança nos resultados apresentados pelo sistema. Dessa forma, o trabalho se insere na agenda de pesquisa em SI sob uma perspectiva sociotécnica [Nunes et al. 2017].

Este trabalho faz parte de um projeto mais abrangente, em desenvolvimento no Laboratório de Aplicações de Informática em Saúde (LApIS), da Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP), em parceria com o Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (InCor/HCFMUSP), que visa a desenvolver um sistema de auxílio ao diagnóstico de cardiomiopatias. Esse projeto inclui etapas de segmentação do VE [Ribeiro 2022, Ribeiro and Nunes 2023, Ribeiro and Nunes 2022a], extração de características [Bergamasco et al. 2018], recuperação baseada em conteúdo [Bergamasco et al. 2022] e classificação [Gonçalves and Nunes 2021].

O trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos correlatos sobre detecção e correção de erros; a Seção 3 descreve os materiais e métodos, incluindo o sistema ao qual o método foi integrado e o protocolo experimental; a Seção 4 apresenta e discute os resultados; e a Seção 5 apresenta as conclusões.

2. Trabalhos correlatos

A segmentação de estruturas cardíacas em imagens médicas tem sido amplamente estudada, com métodos que vão desde abordagens clássicas, como modelos deformáveis, até técnicas baseadas em inteligência artificial [Ribeiro and Nunes 2022b]. No entanto, embora a presença de erros em segmentações seja comum, poucos estudos abordam explicitamente a detecção e correção desses erros. Para compreender o que tem sido desenvolvido na literatura, a revisão dos trabalhos correlatos foi organizada em duas subseções. A primeira aborda a detecção de erros em segmentações (Seção 2.1), enquanto a segunda discute técnicas de correção (Seção 2.2).

2.1. Abordagens para detecção de erros em segmentações

No que tange à avaliação da qualidade de segmentações, uma linha de métodos baseia-se no conceito de *Reverse Classification Accuracy* (RCA) [Valindria et al. 2017]. A ideia central do RCA é treinar um segundo classificador a partir da segmentação prevista e avaliar seu desempenho em um banco de imagens com GT — bom desempenho desse classificador sugere provável qualidade da segmentação, pois o método parte do pressuposto

de que uma segmentação precisa fornece rótulos consistentes que permitem ao classificador aprender características discriminativas que generalizam para imagens anotadas por especialistas; por outro lado, segmentações imprecisas ou inconsistentes tendem a gerar rótulos ruidosos, o que resulta em pior desempenho do classificador frente ao GT. Diversos estudos têm verificado a aplicabilidade e a validação do RCA em conjuntos de dados extensos [Robinson et al. 2019, Robinson et al. 2017]. Mais recentemente, o *In-Context* RCA foi proposto para incorporar informação de contexto e reduzir a dependência de um grande banco com segmentações do especialista [Cosarinsky et al. 2025].

Ademais, outra abordagem recorre às características extraídas das imagens segmentadas, os *radiomics*, para treinar modelos de controle de qualidade [Liu et al. 2023]. Esses *pipelines* extraem um amplo conjunto de características (texturais, morfológicas, de intensidade) e empregam classificadores ou modelos de regressão para identificar segmentações falhas [Izquierdo et al. 2021]. O estudo de [Sander et al. 2020] combina segmentação automática com mapas de incerteza para treinar uma rede capaz de detectar erros em RMC. Revisões recentes enfatizam que estimativas de incerteza são relevantes para a detecção, e discutem técnicas como aproximações bayesianas e calibração de probabilidades para indicar erros em segmentações [Zou et al. 2023, Abdar et al. 2021]. Por fim, [Arega et al. 2025] propõe uma abordagem pós-hoc para detecção de segmentações fora da distribuição, baseada na distância entre as características extraídas pelo *encoder* durante a segmentação e o vetor médio de características das segmentações de validação: quando essa distância excede um limiar, a segmentação é considerada fora da distribuição.

Vale destacar duas limitações importantes desses métodos: i) a maioria foi concebida para detectar erros em fatias isoladas e não aborda explicitamente erros sequenciais ao longo da série temporal — enquanto o presente trabalho focaliza precisamente erros sequenciais; e ii) muitos desses métodos, em especial o RCA, implicam elevado custo computacional e dependem de bases anotadas para treinamento, o que restringe sua aplicabilidade em cenários com recursos de anotação limitados. Tais limitações motivam a investigação de estratégias capazes de incorporar informação temporal e reduzir a necessidade de grandes conjuntos rotulados.

2.2. Técnicas de correção de segmentações

Diversas abordagens utilizam modelos deformáveis com o objetivo de aprimorar os contornos inicialmente estimados por métodos automáticos. Em trabalhos que integram redes de AP com modelos deformáveis, uma segmentação inicial produzida por AP é refinada ao minimizar uma funcional de energia que incorpora restrições de forma e características específicas do exame, resultando em contornos mais plausíveis e consistentes [Avendi et al. 2016, Ribeiro and Nunes 2023, Yang et al. 2013, Ribeiro and Nunes 2022b]. Algumas abordagens exploram modelos direcionados à modelagem das bordas por meio de estruturas probabilísticas, como campos aleatórios condicionais, que integram aprendizado de bordas e informações de forma para impulsionar a precisão local do contorno, especialmente em regiões com baixo contraste ou presença de músculos papilares [Dreijer et al. 2013]. Paralelamente, existem métodos que garantem propriedades anatômicas, incluindo topologia e a morfologia do ventrículo, durante o processo de segmentação [Painchaud et al. 2020]. Contudo, essas abordagens frequentemente exigem modelagens complexas, alto custo computacional e podem impor restrições que não generalizam bem em toda a variabilidade anatômica.

A proposta deste trabalho é realizar um comparativo entre diferentes métodos de interpolação temporal, justificando-os pelo baixo custo computacional e pela facilidade de integração em *pipelines* clínicos em comparação com soluções mais complexas. Além disso, não é comum na literatura considerar explicitamente informações das imagens vizinhas na sequência e os métodos costumam depender da formulação e calibração cuidadosa de funcionais de energia, o que torna sua aplicação prática mais trabalhosa. Os trabalhos de Juhl et al. [Juhl et al. 2024], Aganj et al. [Aganj et al. 2017], Mortensen e Taylor [Mortensen and Taylor 1999], e Raya e Udupa [Raya and Udupa 1990] serviram como base conceitual para o desenvolvimento das abordagens de correção propostas neste estudo. Esses métodos exploram diferentes formas de representar e deformar estruturas anatômicas — desde o uso de campos de distância assinados e modelagem estatística de forma, até registros deformáveis e interpolação morfológica ou baseada em forma.

3. Materiais e métodos

Como mencionado na Seção 1, o objetivo principal deste trabalho é desenvolver um método de pós-processamento acoplado após a etapa de segmentação em sistemas CAD de cardiopatias, para identificar e corrigir erros em sequências temporais de segmentações do VE em exames médicos, visando aprimorar a qualidade dos resultados e ampliar sua utilidade em sistemas de apoio à decisão clínica. A Seção 3.1 descreve o sistema que o método de pós-processamento proposto será integrado.

O desenvolvimento do trabalho foi estruturado em quatro etapas principais: (i) aquisição das segmentações, contemplando a seleção e organização dos exames utilizados (Seção 3.2); (ii) desenvolvimento do método de detecção, responsável por identificar falhas nas segmentações (Seção 3.3); (iii) desenvolvimento do método de correção, voltado a ajustar os erros preservando a consistência temporal (Seção 3.4); e (iv) validação do método, na qual foram conduzidos experimentos para avaliar a eficácia da abordagem proposta (Seção 3.5).

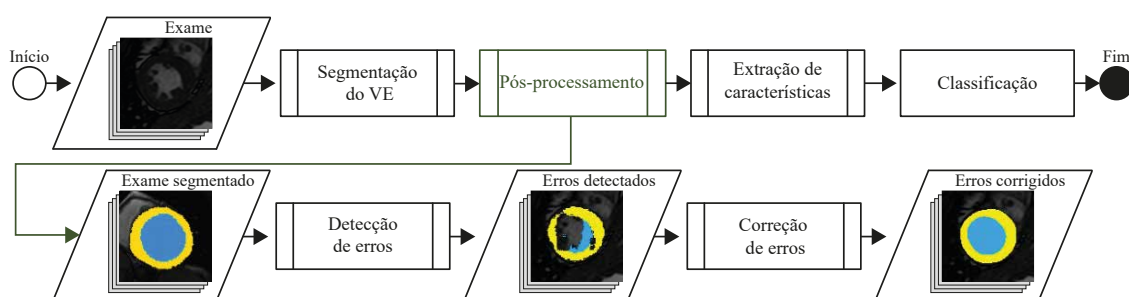
3.1. Sistema CAD para diagnóstico de cardiomiopatias

Um CAD é composto por uma cadeia de etapas interdependentes, que incluem a aquisição das imagens médicas, a segmentação das estruturas anatômicas de interesse, a extração de biomarcadores quantitativos e, por fim, os resultados da classificação a partir dos biomarcadores. Nesse fluxo, a segmentação assume um papel central, pois fornece as bases sobre as quais os indicadores clínicos são calculados. Erros nessa fase podem propagar-se ao longo do processo, comprometendo a qualidade dos biomarcadores extraídos e, potencialmente, modificando os resultados apresentados ao especialista. Desse modo, o método de pós-processamento proposto visa diminuir a incidência de erros na segmentação final produzida pelo sistema, aumentando a confiabilidade dos biomarcadores extraídos e das classificações realizadas nas etapas subsequentes.

Este trabalho insere-se em um projeto em desenvolvimento no Laboratório de Aplicações de Informática em Saúde (LApIS, EACH-USP) em parceria com o Instituto do Coração do Hospital das Clínicas da Faculdade de Medicina da USP (InCor/HCFMUSP), que abrange segmentação do VE [Ribeiro 2022, Ribeiro and Nunes 2023, Ribeiro and Nunes 2022a], extração de características [Bergamasco et al. 2018], recuperação baseada em conteúdo

[Bergamasco et al. 2022] e classificação [Gonçalves and Nunes 2021, Costa et al. 2025]. A Figura 1 mostra a visão geral do sistema CAD para diagnóstico de cardiomiopatias em desenvolvimento, incluindo a nova etapa de pós-processamento proposta.

Figura 1. Visão geral do sistema CAD em desenvolvimento incluindo a etapa de pós-processamento proposta (destacada em verde). O processo recebe como entrada um exame contendo imagens para diferentes fatias e instantes de tempo, sendo que o miocárdio (em amarelo) e a câmara cardíaca (em azul) são então segmentados pelo método de segmentação. A etapa de pós-processamento proposta verifica as segmentações para detectar inconsistências sequenciais e aplica correções nos casos detectados. Por fim, é realizada a extração de características e a classificação - também entendida como diagnóstico automático.



3.2. Aquisição das segmentações

Neste trabalho, foi utilizado o *dataset* LVQuan19¹, proveniente do *Left Ventricle Full Quantification Challenge*, evento da comunidade *The Medical Image Computing and Computer Assisted Intervention Society* de 2019. Esse conjunto é composto por 1120 imagens de RMC em eixo curto, correspondentes a 56 pacientes, cada um com 20 quadros cobrindo todo o ciclo cardíaco, com imagens adquiridas com sincronização por eletrocardiograma [Xue et al. 2021]. Dessa forma, o último e o primeiro quadro da sequência são sequenciais. Todas as imagens possuem a segmentação do especialista, característica que favorece a análise temporal das segmentações e que motivou a escolha desse *dataset*.

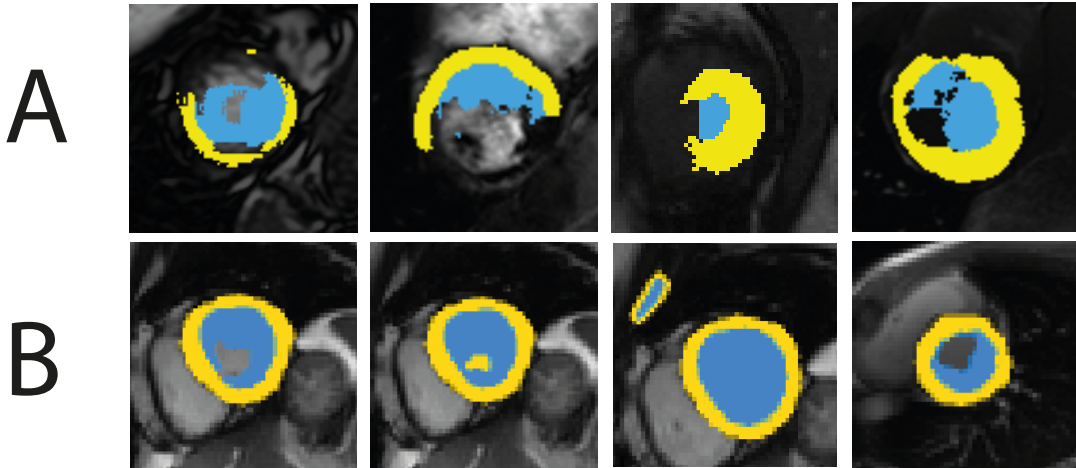
A partir dos exames selecionados, foram consideradas duas formas distintas de obtenção de segmentações problemáticas. A primeira está relacionada a erros reais, resultantes da aplicação de um método de AP composto por três etapas principais: (1) a obtenção automática da região de interesse (ROI); (2) o pré-processamento da ROI; e (3) a aplicação da rede de segmentação U-Net, responsável por classificar os *pixels* entre câmara cardíaca, miocárdio e fundo. O método completo encontra-se descrito em [Ribeiro and Nunes 2021]. As segmentações resultantes foram inspecionadas individualmente, e cada instante de tempo foi categorizado manualmente como correto ou com erro, de modo a fornecer a referência necessária para a avaliação do trabalho. A Figura 2A ilustra exemplos de erros reais gerados pelo método de AP utilizado.

A segunda forma de aquisição de segmentações problemáticas consistiu na geração artificial de erros controlados, por meio do arcabouço proposto em [Raquel et al. 2025]. Esse arcabouço possibilita a inserção de diferentes tipos de falhas —

¹<https://lvquan19.github.io/>

como vazamentos de contornos, buracos e deslocamentos estruturais — além da regulação parametrizável da intensidade do erro. Dessa forma, cria-se um ambiente controlado que favorece a investigação do desempenho do método proposto frente a variações específicas e graduais nos erros, sendo que esse conjunto de dados auxiliou no desenvolvimento do método proposto. A Figura 2B ilustra exemplos de erros simulados.

Figura 2. Exemplo de segmentações do conjunto de dados. (A) Exemplos reais gerados por redes de AP; (B) exemplos sintéticos gerados com o arcabouço proposto em [Raquel et al. 2025].



3.3. Método de detecção

A estratégia de detecção de erros sequenciais parte da premissa de que uma sequência temporal correta apresenta variações suaves entre quadros adjacentes, enquanto falhas de segmentação provocam saltos ou discrepâncias locais nas medidas que descrevem a forma ou a posição das estruturas. Portanto, considera-se que erros sequenciais possam ser detectados a partir da análise de métricas de similaridade.

Para cada sequência temporal considerada, métricas de similaridade definidas pela função $f(t, t + 1)$ são calculadas entre as segmentações de dois instantes de tempo consecutivos t e $t + 1$. Essa função é avaliada para todos os pares adjacentes ao longo do ciclo, com $t = 1, \dots, T$, sendo T o número total de quadros da sequência. No caso especial em que $t = T$, considera-se $t + 1 = 1$, de modo que $f(T, 1)$ representa a comparação entre o último e o primeiro quadro — o que é válido porque os exames utilizados cobrem um ciclo cardíaco completo, conforme descrito na Seção 3.2. A partir dos valores obtidos para todos os pares $(t, t + 1)$, computam-se a mediana m e o desvio padrão σ . Um instante t é marcado como potencialmente incorreto quando o valor de $f(t, t + 1)$ fica fora do intervalo definido por m e σ , segundo o parâmetro de sensibilidade $\alpha > 0$, que regula a tolerância à variabilidade natural (Equação 1).

$$m - \alpha \sigma < f(t, t + 1) < m + \alpha \sigma, \quad (1)$$

O problema pode ser formalizado como uma tarefa de classificação binária, cujo objetivo é atribuir a cada instante de tempo t uma variável indicadora $y_t \in \{0, 1\}$, onde a

classe positiva $y_t = 1$ denota a presença de erro na segmentação correspondente e $y_t = 0$ indica uma segmentação considerada sem erros. A decisão sobre o rótulo y_t é baseada nas métricas de similaridade $f(t, t + 1)$ calculadas entre instantes consecutivos, as quais quantificam o grau de continuidade temporal entre as segmentações. Dessa forma, o classificador utiliza o comportamento dessas métricas ao longo do ciclo — em relação à mediana m , ao desvio padrão σ e ao parâmetro de sensibilidade α — para identificar automaticamente instantes anômalos no exame. Essa formalização pode ser resumida pela Equação 2

$$y_t = \begin{cases} 1, & \text{se } f(t, t + 1) \notin [m - \alpha\sigma, m + \alpha\sigma], \\ 0, & \text{caso contrário.} \end{cases} \quad (2)$$

3.3.1. Métricas de similaridade avaliadas

Foram implementadas e testadas métricas que avaliam a continuidade temporal das segmentações. A seguir, descreve-se o significado e a forma de cálculo de cada uma delas, assim como decisões de implementação relevantes.

Coefficiente Dice. Para cada classe anatômica (miocárdio e câmara cardíaca) calcula-se o coeficiente de sobreposição Dice, definido pela Equação 3, onde S_t e S_{t+1} são as máscaras binárias nos dois instantes consecutivos. Além das duas medidas por classe, é reportada a média simples entre o Dice do miocárdio e o Dice da câmara cardíaca.

$$\text{Dice}(t, t + 1) = \frac{2|S_t \cap S_{t+1}|}{|S_t| + |S_{t+1}|}, \quad (3)$$

Distância de Hausdorff. A discrepância de contorno é quantificada pela distância de Hausdorff simétrica, definida pela Equação 4, onde $d(\cdot, \cdot)$ é a distância euclidiana. No contexto do trabalho, os conjuntos S_t e S_{t+1} correspondem às coordenadas dos pontos de contorno das segmentações nos instantes t e $t + 1$, respectivamente. Calcula-se as distâncias em ambas as direções e toma-se o máximo como Hausdorff simétrico. Essa métrica é computada separadamente para o miocárdio, para a câmara cardíaca e considerando ambas as estruturas conjuntamente.

$$H(t, t + 1) = \max \left\{ \sup_{s_t \in S_t} \inf_{s_{t+1} \in S_{t+1}} d(s_t, s_{t+1}), \sup_{s_{t+1} \in S_{t+1}} \inf_{s_t \in S_t} d(s_{t+1}, s_t) \right\}. \quad (4)$$

Varição de área (absoluta e proporcional). As áreas segmentadas são obtidas pela quantidade de pixels pertencentes a cada uma das máscaras, miocárdio e câmara, para cada instante. Dadas as duas áreas A_t e A_{t+1} , calcula-se a variação absoluta (Equação 5) e a variação proporcional (Equação 6).

$$\Delta A = |A_{t+1} - A_t|, \quad (5)$$

$$\Delta A_{\text{prop}} = \left| 1 - \frac{A_t}{A_{t+1}} \right|, \quad (6)$$

Varição da assinatura. Para capturar alterações locais na forma, extrai-se a assinatura radial da forma a partir do centróide da máscara combinada (união de miocárdio e câmara cardíaca), seguindo a descrição de Gonzalez & Woods [Rafael C. Gonzalez 2002]. A assinatura consiste no vetor de distâncias euclidianas do centróide até os pontos referentes ao epicárdio. Calcula-se o desvio padrão dessas distâncias como medida de irregularidade da forma. A variação temporal da assinatura entre dois instantes é então definida pela diferença absoluta entre os desvios padrão - como descreve a Equação 7).

$$\Delta_{\text{sig}} = |\sigma_{t+1} - \sigma_t|, \quad (7)$$

3.3.2. Detecção de erros com uma única métrica

A primeira etapa experimental consistiu em avaliar a detecção considerando isoladamente cada métrica. Para cada métrica testou-se um conjunto amplo de limiares, representado pelo parâmetro α , suficiente para observar comportamentos sob diferentes sensibilidades. A escolha do melhor limiar para uma métrica individual foi realizada mediante avaliação sobre dados rotulados reais usando medidas de desempenho de detecção, por exemplo, F_1 -score (Seção 3.5). Esse procedimento torna possível estabelecer, de modo empírico, uma configuração que equilibre falsos positivos e falsos negativos para cada métrica isolada. Ademais, para cada métrica, as versões calculadas para o miocárdio, para a câmara cardíaca e ambas conjuntamente foram tratadas como métricas distintas; isto é, limiares e avaliações foram determinados separadamente para cada classe, possibilitando capturar sensibilidade diferencial dos critérios às falhas em cada estrutura.

3.3.3. Detecção de erros com combinação de métricas

Para investigar se a combinação de métricas melhora a detecção, procedeu-se da seguinte forma: para cada métrica foi retido o melhor limiar (segundo F_1) obtido na etapa anterior. Em seguida, foram geradas combinações entre esses limiares de métricas distintas e testadas regras de fusão simples: a união lógica e intersecção. A *união lógica (OR)* refere-se a detectar erro sempre que ao menos uma métrica indicar erro (Equação 8); intersecção (*AND*), por sua vez, detecta erro apenas quando todas as métricas concordam (Equação 9). Cada combinação foi avaliada contra a mesma referência manual, tornando possível comparar desempenho entre estratégias individuais e combinadas. O objetivo foi verificar se combinações judiciais de métricas reduzem falsos positivos e/ou aumentam a sensibilidade na detecção de erros sequenciais.

$$E_{\text{OR}} = E_1 \vee E_2 \vee \dots \vee E_n, \quad (8)$$

$$E_{\text{AND}} = E_1 \wedge E_2 \wedge \dots \wedge E_n. \quad (9)$$

3.4. Método de correção

Após a etapa de detecção, os instantes classificados como problemáticos são encaminhados para a etapa de correção, responsável por restabelecer a consistência da sequência temporal. Neste trabalho, foram considerados quatro algoritmos distintos de correção, com o intuito de realizar uma comparação entre seus resultados. Todos os métodos recebem como entrada duas segmentações de instantes de tempo vizinhos - uma anterior, S_{t-1} , e outra posterior, S_{t+1} , ao quadro considerado problemático, S_t — e, a partir dessas referências, produzem uma nova segmentação corrigida, \hat{S}_t .

Ademais, todos os métodos aqui descritos fazem uso de interpolação como princípio de correção. Três deles empregam a representação baseada em Signed Distance Functions (SDF), enquanto apenas um método opera diretamente nas coordenadas dos pontos dos contornos. O SDF atribui a cada ponto do domínio a menor distância ao contorno da região segmentada com um sinal que indica interior e exterior [Osher and Fedkiw 2003]. Formalmente, para uma região Ω com contorno $\partial\Omega$, a SDF é definida pela Equação (10). Vale destacar que, após a execução dos algoritmos de correção, aplica-se uma etapa destinada a garantir a ausência de interseções entre classes. Dessa forma, qualquer sobreposição entre câmara e miocárdio é corrigida. Os algoritmos de correção serão descritos em detalhe a seguir.

$$\phi(x) = \begin{cases} -\min_{y \in \partial\Omega} \|x - y\|, & x \in \Omega, \\ \min_{y \in \partial\Omega} \|x - y\|, & x \notin \Omega, \end{cases} \quad (10)$$

3.4.1. Interpolação de contornos

Diferentemente dos demais métodos que operam no espaço das SDF, a interpolação de contornos atua diretamente sobre as coordenadas dos pontos que descrevem o contorno das segmentações das estruturas do VE. A partir das segmentações de entrada do método de detecção, o método extrai os contornos das estruturas anatômicas de interesse. Em seguida, realiza uma reamostragem desses contornos, de modo que ambos sejam representados por um mesmo número de pontos igualmente distribuídos entre o comprimento do contorno. A Equação 11 considera C_1 e C_2 como os conjuntos de pontos pertencentes aos contornos dos instantes vizinhos, em que cada ponto é ordenado de acordo com sua posição proporcional ao comprimento total do contorno. Desse modo, os pontos $C_1[i]$ e $C_2[i]$ correspondem à mesma fração do percurso do contorno em relação ao ponto inicial. A interpolação é então conduzida ponto a ponto, a partir da Equação (12), utilizando um fator de ponderação $\alpha \in [0, 1]$ para controlar a contribuição relativa de cada instante e $C_{\text{interp}}(\alpha)[i]$ corresponde ao ponto i do contorno interpolado.

$$C_1 = \{\mathbf{c}_{1,i}\}_{i=0}^{n-1}, \quad C_2 = \{\mathbf{c}_{2,i}\}_{i=0}^{n-1} \quad (11)$$

$$\mathbf{C}_{\text{interp}}(\alpha)[i] = \alpha \mathbf{c}_{1,i} + (1 - \alpha) \mathbf{c}_{2,i}, \quad i = 0, \dots, n - 1, \quad \alpha \in [0, 1], \quad (12)$$

A partir desse novo contorno, é construída uma máscara preenchida que representa a estimativa corrigida da estrutura. Esse processo é aplicado separadamente para cada classe anatômica, garantindo que a hierarquia entre as regiões seja preservada.

3.4.2. Interpolação via SDF

A estratégia denominada interpolação via SDF gera uma estimativa intermediária a partir de duas segmentações vizinhas operando no domínio das SDFs. O procedimento inicia-se pela conversão das máscaras binárias das classes de interesse nas correspondentes SDFs em cada instante de referência. Para cada classe anatômica relevante, individualmente, calcula-se então uma SDF intermediária por interpolação, descrito pela Equação(13) onde $\phi^{(1)}$ e $\phi^{(2)}$ são as SDFs nos instantes anterior e posterior, respectivamente, e $\alpha \in [0, 1]$ regula a contribuição relativa de cada referência (o caso $\alpha = 0.5$ corresponde ao ponto médio). O método finaliza com a combinação das máscaras.

$$\phi^{\text{mid}}(x) = \alpha \phi^{(1)}(x) + (1 - \alpha) \phi^{(2)}(x), \quad (13)$$

3.4.3. Interpolação por registro deformável sobre SDF

A estratégia de interpolação por registro deformável (RD) é guiado pela representação em SDF da união das estruturas anatômicas de interesse, de modo que o registro empenha-se em alinhar os contornos das formas entre os dois instantes temporais adjacentes passados como entrada. Antes do registro, as SDFs de referência são suavizadas por um filtro Gaussiano para atenuar ruídos locais e favorecer um campo de deformação mais regular.

O núcleo do método consiste em estimar um campo de deslocamento que mapeia a SDF do instante móvel para a SDF do instante fixo, a partir de um algoritmo clássico de RD. O algoritmo Demons foi proposto por [Thirion 1998] e utilizado neste trabalho para o RD. O resultado é um campo vetorial contínuo $\mathbf{u}(x)$ que descreve a transformação necessária para alinhar as duas representações de distância. Esse campo contém direção e magnitude em cada ponto; portanto, escalá-lo por um fator de ponderação altera a magnitude dos vetores, isto é, a distância que cada ponto deve percorrer, mantendo suas direções. Escalar \mathbf{u} por α produz uma deformação parcial proporcional a α , de modo que $\alpha = 0.5$ move cada ponto metade do deslocamento originalmente estimado, conforme a Equação 14. A máscara corrigida é então obtida ao aplicar o campo de deslocamento parcial sobre as máscaras de referência, utilizando interpolação do tipo vizinho mais próximo para preservar rótulos discretos.

$$\mathbf{u}_{\text{mid}}(x) = \alpha \mathbf{u}(x), \quad (14)$$

3.4.4. Interpolação no espaço do Modelo Estatístico de Forma (PCA sobre SDF)

O método inicia com a transformação da entrada para SDF, sendo que, para preservar a informação de cada estrutura, o cálculo é realizado separadamente para cada máscara. Após isso, as SDFs são empacotadas por achatamento em um vetor de alta dimensão; concatenando os vetores das diferentes classes obtém-se um único vetor que codifica a geometria completa de cada instante, pois contém a distância de cada ponto da máscara ao limite das estruturas. Esse procedimento converte imagens bidimensionais contínuas em amostras vetoriais - em que cada componente do vetor é a amostra da SDF em uma posição de pixel - viabilizando tratar cada instante como um ponto num conjunto sobre o qual se aplica análise estatística de forma.

Em seguida, constrói-se um Modelo Estatístico de Forma por meio de Análise de Componentes Principais (PCA) aplicada às amostras vetoriais das SDFs. O procedimento inclui: a centralização dos dados em torno da média, de modo a analisar somente variações relativas do SDF em torno dessa média; a estimação da matriz de covariância; e a decomposição espectral para extrair as direções de maior variância (autovetores), que definem a base do espaço latente. Cada amostra é projetada nessa base para obter coordenadas latentes compactas; a interpolação é então realizada nesse espaço reduzido por meio de uma combinação linear entre as coordenadas latentes das duas referências temporais, produzindo uma posição latente intermediária que sintetiza a contribuição de ambas as entradas.

A coordenada interpolada no espaço latente é projetada de volta ao domínio das SDFs pela reconstrução do modelo, produzindo SDFs intermediárias para cada classe. Essas SDFs reconstruídas são convertidas em máscaras binárias por limiarização no nível zero, ou seja, pontos com valores menores ou iguais a zero são considerados pertencentes à região de interesse, formalmente definida pela Equação 15, onde $\phi^{(c)}(x)$ representa o valor da SDF reconstruída para a classe c no ponto x .

$$M^{(c)}(x) = \begin{cases} 1, & \text{se } \phi^{(c)}(x) \leq 0, \\ 0, & \text{caso contrário,} \end{cases} \quad (15)$$

3.5. Experimento de validação

O estudo segue uma abordagem empírico-experimental quantitativa: algoritmos de detecção e correção foram desenvolvidos e avaliados em experimentos controlados sobre o conjunto LVQuan19. O desenho experimental foi estruturado para isolar o efeito de cada falha na cadeia detecção–correção e permitir comparações quantitativas entre estratégias de correção. Para tal, a partir do conjunto de dados reais citado na Seção 3.2, para cada exame, dada uma sequência temporal com X instantes contendo erro, geraram-se X novas sequências: em cada cópia manteve-se somente um dos instantes com erro, o qual constitui o único erro da cópia, e substituíram-se os demais instantes pela segmentação GT correspondente. Para obtenção dos limiares e calibração das métricas a divisão dos dados foi realizada com dados reais estratificados entre treino (80%) e teste (20%), mantendo a proporção entre diferentes intensidades de erro.

A avaliação da detecção foi realizada com a métrica *FI-score*, calculada a par-

tir das métricas de precisão e sensibilidade, sobre a tarefa de identificar instantes problemáticos. Denotando por TP os verdadeiros positivos, FP os falsos positivos e FN os falsos negativos, empregam-se as Equações 16 e 17:

$$precisão = \frac{TP}{TP + FP}, \quad (16)$$

$$sensibilidade = \frac{TP}{TP + FN}. \quad (17)$$

A métrica *F1-score* é dada pela média harmônica entre a *precisão* e a *sensibilidade* (Equação 18), sendo que a medida calculada refere-se ao nível por quadro: cada instante classificado pelo detector é comparado com o rótulo binário “problemático / não-problemático”.

$$F_1 = 2 \cdot \frac{precisão \cdot sensibilidade}{precisão + sensibilidade}. \quad (18)$$

Para avaliar a correção o coeficiente de similaridade Dice foi utilizado, sendo calculado entre a máscara corrigida e a máscara GT correspondente, a partir da fórmula descrita na Seção 3.3.1. Para análise por níveis de severidade, os erros originais, anteriores à correção, foram agrupados segundo o Dice calculado entre a segmentação com erro e o GT: 951 casos leves — Dice superior a 0,80; 182 casos moderados — Dice superior a 0,60 e até 0,80; 242 casos severos — Dice inferior a 0,60. Essas faixas foram utilizadas para comparar desempenho de detecção e eficácia das rotinas de correção em diferentes níveis de severidade.

4. Resultados e discussões

Nesta seção, apresenta-se os resultados do método de pós-processamento desenvolvido. Na Seção 4.1, são discutidos os experimentos de detecção, comparando o desempenho obtido com o uso de métricas individuais e combinações de métricas. Em seguida, a Seção 4.2 aborda os resultados da etapa de correção, analisando os diferentes métodos implementados e demonstrando a eficácia da proposta na correção de erros reais produzidos por modelos de AP.

4.1. Método de detecção

A Tabela 1 e a Tabela 2 mostram que o detector funciona bem em erros moderados e severos, enquanto a identificação de erros leves continua sendo o maior desafio. Em termos numéricos, erros severos são praticamente detectados por qualquer configuração (*F1-score* de 0.99), ao passo que para erros leves o desempenho é inferior: usando uma única métrica o *F1-score* é 0,62, com precisão 0,55 e sensibilidade 0,70. Uma hipótese que justifica essa dificuldade do método em detectar erros leves é porque essa faixa de erro fica próxima da variabilidade anatômica natural entre quadros, resultando em diferenças sutis que permanecem abaixo do limiar de sensibilidade das métricas utilizadas.

A combinação de métricas melhora a detecção em cenários onde a anomalia é pequena: para erros leves o *F1-score* sobe de 0,62 para 0,68 e a precisão aumenta de

0,55 para 0,71, embora haja uma pequena queda na sensibilidade (0,70 para 0,66). Desse modo, a combinação de métricas aumenta a precisão ao exigir concordância entre critérios distintos para sinalizar um erro, o que reduz falsos positivos, mas torna o detector mais conservador e passa a omitir alguns verdadeiros positivos, reduzindo a sensibilidade. Em erros moderados observa-se também ganho (*F1-score* de 0,96 para 0,98; precisão de 0,98 para 0,99; sensibilidade de 0,94 para 0,97), o que indica que métricas complementares consolidam evidências quando a falha gera discrepâncias detectáveis por múltiplas métricas, mesmo que não sejam severas. Em termos operacionais, quando a prioridade for maximizar a sensibilidade, para capturar o máximo de falhas, os resultados sugerem privilegiar limiares mais permissivos ou fusões por *OR*.

Para erros leves, a melhor combinação reuniu a média do Dice entre miocárdio e câmara cardíaca ($\alpha = 0,7$), o Dice do miocárdio ($\alpha = 1,0$) e a média da distância de Hausdorff entre miocárdio e câmara cardíaca ($\alpha = 0,9$). Para erros moderados, a configuração mais eficaz foi composta pela média do Dice entre miocárdio e câmara cardíaca ($\alpha = 0,7$), pela distância de Hausdorff do miocárdio ($\alpha = 0,7$) e pela média da distância de Hausdorff ($\alpha = 1,3$). Em casos severos, diversas métricas apresentaram desempenho similar, com destaque para o Dice e a distância de Hausdorff, calculados por estrutura individual ou pela média entre miocárdio e câmara cardíaca.

Tabela 1. Desempenho de detecção utilizando a melhor métrica individual (Dice do miocárdio).

Severidade do Erro	<i>F1-score</i>	Precisão	Sensibilidade
Leve	0,62	0,55	0,70
Moderado	0,96	0,98	0,94
Severo	0,99	0,99	0,99

Tabela 2. Desempenho de detecção utilizando a melhor combinação de métricas. (Dice do miocárdio, média do Dice entre miocárdio e câmara cardíaca e média do Hausdorff entre miocárdio e câmara cardíaca).

Severidade do Erro	<i>F1-score</i>	Precisão	Sensibilidade
Leve	0,68	0,71	0,66
Moderado	0,98	0,99	0,97
Severo	0,99	0,99	0,99

4.2. Método de correção

A Figura 3 ilustra dois exemplos de correção para cada intensidade de erro, sendo possível visualizar como a correção aumentou o valor da métrica Dice para todos os casos e melhorou a consistência sequencial entre os instantes $t-1$, t e $t+1$. A Tabela 3 evidencia o impacto das diferentes estratégias de correção sobre os casos de segmentação com erros de diferentes intensidades (leve, moderado e severo). Na condição base, Tabela 3A, sem correção, observa-se uma queda acentuada nas métricas à medida que a severidade do erro aumenta: o Dice médio varia de 0,87 (leve) para 0,33 (severo), revelando que a

degradação da segmentação compromete fortemente a métrica Dice para o miocárdio e câmara cardíaca.

Após a aplicação das estratégias de correção, verifica-se uma melhora substancial no Dice em todos os métodos. A melhoria no Dice médio indica segmentações mais próximas ao contorno de referência, o que reduz erros na extração de biomarcadores clínicos derivados dessas máscaras, ou seja, ao diminuir discrepâncias estruturais na segmentação, reduz-se também a variação indesejada nas medidas clínicas subsequentes, tornando os indicadores mais confiáveis. A interpolação de contornos, Tabela 3B, e a interpolação via SDF, Tabela 3C, apresentaram desempenhos bastante semelhantes, atingindo Dice médio próximo de 0,95 mesmo em erros severos. Esses resultados sugerem que ambas as abordagens são capazes de reconstruir os contornos perdidos ou distorcidos de forma robusta, inclusive em situações em que a segmentação inicial estava altamente comprometida. Já a interpolação por RD sobre SDF, Tabela 3D, apresentou resultados ligeiramente inferiores, com Dice médio variando entre 0,90 e 0,92, possivelmente porque o RD é sensível a pequenas diferenças locais como desalinhamentos entre instantes temporais que podem gerar campos de deslocamento imprecisos.

Por fim, a interpolação com PCA sobre SDF, Tabela 3E, mostrou os melhores resultados gerais, alcançando Dice médio de até 0,96. O desempenho superior do PCA pode estar associado à sua capacidade de capturar as principais componentes da variação da forma, reconstruindo de maneira mais fiel as estruturas cardíacas mesmo diante de movimentos mais intensos do VE, como grandes contrações ou deslocamentos entre os instantes do ciclo cardíaco. Em outras palavras, enquanto os demais métodos dependem fortemente da interpolação local, o PCA atua sobre a forma como um todo, o que pode favorecer melhores resultados.

Tabela 3. Resultados do método de correção.

(a) Erros não corrigidos (casos base)

Erro	Dice Mio	Dice Endo	Dice Médio
Leve	0,81	0,92	0,87
Moderado	0,65	0,82	0,73
Severo	0,40	0,26	0,33

(b) Interpolação de contornos

Erro	Dice Mio	Dice Endo	Dice Médio
Leve	0,93	0,97	0,95
Moderado	0,89	0,97	0,93
Severo	0,92	0,97	0,94

(c) Interpolação via SDF

Erro	Dice Mio	Dice Endo	Dice Médio
Leve	0,93	0,97	0,95
Moderado	0,92	0,97	0,94
Severo	0,93	0,97	0,95

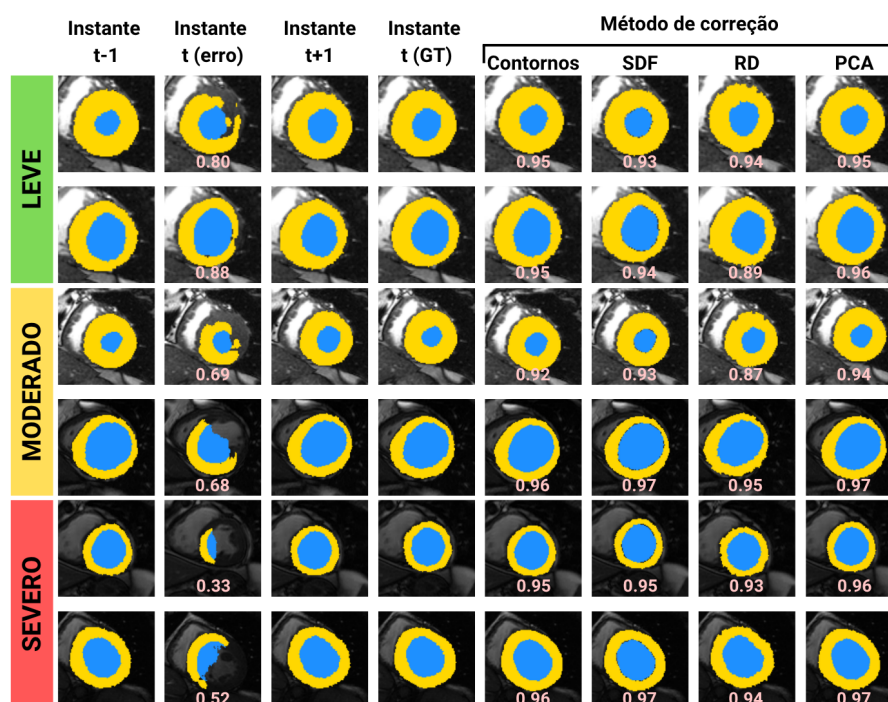
(d) Interpolação por RD sobre SDF

Erro	Dice Mio	Dice Endo	Dice Médio
Leve	0,88	0,95	0,91
Moderado	0,86	0,95	0,90
Severo	0,89	0,96	0,92

(e) Interpolação com PCA sobre SDF

Erro	Dice Mio	Dice Endo	Dice Médio
Leve	0,93	0,97	0,96
Moderado	0,92	0,95	0,93
Severo	0,93	0,97	0,96

Figura 3. Exemplos de correção de erros em segmentações do VE para diferentes intensidades (leve, moderado e severo). Cada linha mostra a segmentação do instante problemático, segmentações dos instantes vizinhos, a segmentação do especialista e a segmentação resultante dos métodos de correção, com valores de Dice calculados em relação a segmentação do especialista.



5. Conclusão

Este trabalho apresentou um método de pós-processamento para segmentações do VE em exames de RMC, composto por uma etapa de detecção de instantes problemáticos seguida por uma etapa de correção baseada em interpolação entre segmentações adjacentes. Os experimentos demonstraram que a abordagem é capaz de identificar erros na sequência temporal e, uma vez detectados, recuperar a consistência sequencial das segmentações, aproximando-as das segmentações de referência. A estratégia de combinação de métricas mostrou-se vantajosa para reduzir a quantidade de falsos positivos, enquanto estratégias com modelos de forma global (PCA sobre SDF) tendem a produzir correções mais robustas frente a erros intensos. O método de detecção obteve bom desempenho, especialmente para casos moderados e severos; o método de correção conseguiu corrigir o erro a partir de segmentações vizinhas, alcançando resultados próximos a segmentação do especialista. O estudo indica que a incorporação experimental desta etapa no sistema CAD é vantajosa e pode ajudar a melhorar estimativas e suportar a decisão clínica.

Embora desenvolvido e avaliado no contexto da cardiologia — especificamente para o pós-processamento de segmentações do ventrículo esquerdo em RMC — o método proposto baseia-se em princípios gerais de coerência temporal e medidas de similaridade entre segmentações adjacentes. Ao operar sobre a continuidade temporal das máscaras e por meio de interpolação entre quadros vizinhos, a abordagem é potencialmente aplicável a outros domínios com séries temporais de segmentações, como veias cerebrais [Phellan et al. 2018] e regiões pulmonares [Tavares et al. 2010]. Dessa forma,

o método configura-se como uma ferramenta genérica e reutilizável para a detecção e correção de inconsistências temporais em segmentações ao longo do tempo.

Apesar dos resultados promissores, existem limitações relevantes que devem ser enfatizadas. Primeiro, a avaliação foi conduzida em cenários controlados nos quais a sequência contém apenas um erro por avaliação - uma simplificação que pode não refletir situações clínicas reais e que podem ocorrer ao longo de vários instantes. Segundo, o método de correção pressupõe que os instantâneos vizinhos estão corretos; quando essa hipótese falha, a interpolação pode propagar ou até agravar o erro. Terceiro, a detecção de erros leves permanece um desafio, pois esses casos exigem sensibilidade fina sem aumentar excessivamente falsos positivos.

Como trabalhos futuros, propõe-se: (i) ampliar e diversificar os sinais usados na detecção; (ii) explorar estratégias de correção capazes de lidar simultaneamente com múltiplas falhas ao longo da sequência; (iii) investigar a integração de restrições anatômicas já na etapa de segmentação. Mais informações, bem como a disposição dos artefatos, podem ser acessadas no repositório².

Agradecimentos

Este trabalho foi apoiado, em parte, pelo Instituto Nacional de Ciência e Tecnologia – Medicina Assistida por Computação Científica (INCT-MACC) [processo 2014/50889-7], pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [processos 307710/2022-0 e 122803/2024-8], pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) [processo 2021/14902-2] e pela Pró-Reitoria de Pesquisa e Inovação da USP.

Referências

- Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., and Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297.
- Aganj, I., Iglesias, J. E., Reuter, M., Sabuncu, M. R., and Fischl, B. (2017). Mid-space-independent deformable image registration. *NeuroImage*, 152:158–170.
- Anari, S., Tataei Sarshar, N., Mahjoori, N., Dorosti, S., and Rezaie, A. (2022). Review of deep learning approaches for thyroid cancer diagnosis. *Mathematical Problems in Engineering*, 2022:1–8.
- Arega, T. W., Bricq, S., and Meriaudeau, F. (2025). Post-hoc out-of-distribution detection for cardiac MRI segmentation. *Computerized Medical Imaging and Graphics*, 119:102476.
- Asiri, N., Hussain, M., Al Adel, F., and Alzaidi, N. (2019). Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. *Artificial Intelligence in Medicine*, 99:101701.
- Avendi, M., Kheradvar, A., and Jafarkhani, H. (2016). A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Medical Image Analysis*, 30:108–119.

²https://github.com/Bruno-Friedrich/LApIS-Deteccao_Correcao_Erros

- Bergamasco, L. C. C., Lima, K. R. P. S., Rochitte, C. E., and Nunes, F. L. S. (2018). 3D medical objects retrieval approach using spharms descriptor and network flow as similarity measure. In *Proceedings of the 31st Conference on Graphics, Patterns and Images (SIBGRAPI 2018)*, pages 329–336, Piscataway, NJ, USA. IEEE.
- Bergamasco, L. C. C., Lima, K. R. P. S., Rochitte, C. E., and Nunes, F. L. S. (2022). A bipartite graph approach to retrieve similar 3d models with different resolution and types of cardiomyopathies. *Expert Systems with Applications*, 193:116422.
- Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.-A., Cetin, I., Lekadir, K., Camara, O., Ballester, M. A. G., Sanroma, G., Napel, S., Petersen, S., Tziritas, G., Grinias, E., Khened, M., Kollerathu, V. A., Krishnamurthi, G., Rohe, M.-M., Pennec, X., Sermesant, M., Isensee, F., Jager, P., Maier-Hein, K. H., Full, P. M., Wolf, I., Engelhardt, S., Baumgartner, C. F., Koch, L. M., Wolterink, J. M., Isgum, I., Jang, Y., Hong, Y., Patravali, J., Jain, S., Humbert, O., and Jodoin, P.-M. (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525.
- Cosarinsky, M., Billot, R., Mansilla, L., Jimenez, G., Gaggió, N., Fu, G., Tirer, T., and Ferrante, E. (2025). Conformal in-context reverse classification accuracy: Efficient estimation of segmentation quality with statistical guarantees.
- Costa, S. S. H., Gonçalves, V. M., Ribeiro, M. A. O., and Nunes, F. L. S. (2025). Generalization of cardiomyopathy classification models based on feature descriptors from magnetic resonance imaging. In *Anais do XXV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2025)*, SBCAS 2025, page 943–954. Sociedade Brasileira de Computação - SBC.
- Dreijer, J. F., Herbst, B. M., and du Preez, J. A. (2013). Left ventricular segmentation from MRI datasets with edge modelling conditional random fields. *BMC Medical Imaging*, 13(1).
- Gonçalves, V. M. and Nunes, F. L. S. (2021). Abordagem híbrida para auxílio ao diagnóstico de cardiomiopatias. Projeto de Pesquisa (Doutorado em Ciências) – Programa de Pós-graduação em Sistemas de Informação (PPgSI), Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP).
- Guertari, R., Ayari, H., and Sakly, H. (2023). Computer-aided diagnosis systems: a comparative study of classical machine learning versus deep learning-based approaches. *Knowledge and Information Systems*, 65(10):3881–3921.
- Izquierdo, C., Casas, G., Martin-Isla, C., Campello, V. M., Guala, A., Gkontra, P., Rodríguez-Palomares, J. F., and Lekadir, K. (2021). Radiomics-based classification of left ventricular non-compaction, hypertrophic cardiomyopathy, and dilated cardiomyopathy in cardiovascular magnetic resonance. *Frontiers in Cardiovascular Medicine*, 8.
- Juhl, K. A., Slipsager, J., de Backer, O., Kofoed, K., Camara, O., and Paulsen, R. (2024). Signed distance field based segmentation and statistical shape modelling of the left atrial appendage.
- Liu, Q., Lu, Q., Chai, Y., Tao, Z., Wu, Q., Jiang, M., and Pu, J. (2023). Radiomics-based quality control system for automatic cardiac segmentation: A feasibility study.

Bioengineering, 10(7):791.

- Moon, W. K., Lee, Y.-W., Ke, H.-H., Lee, S. H., Huang, C.-S., and Chang, R.-F. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer Methods and Programs in Biomedicine*, 190:105361.
- Mortensen, E. and Taylor, D. (1999). An image space algorithm for morphological contour interpolation.
- Nunes, V., Cappelli, C., and Ralha, C. (2017). Transparency in information systems. In *I GranDSI-BR: Grandes Desafios da Pesquisa em Sistemas de Informação no Brasil para o período de 2016 a 2026*, pages 73–89. SBC.
- Osher, S. and Fedkiw, R. (2003). *Signed Distance Functions*, page 17–22. Springer New York.
- Painchaud, N., Skandarani, Y., Judge, T., Bernard, O., Lalande, A., and Jodoin, P.-M. (2020). Cardiac segmentation with strong anatomical guarantees. *IEEE Transactions on Medical Imaging*, 39(11):3703–3713.
- Phellan, R., Lindner, T., Helle, M., Falcao, A. X., and Forkert, N. D. (2018). Automatic temporal segmentation of vessels of the brain using 4d asl mra images. *IEEE Transactions on Biomedical Engineering*, 65(7):1486–1494.
- Rafael C. Gonzalez, R. E. W. (2002). *Digital Image Processing*. Prentice Hall, New Jersey, USA, 2nd ed edition.
- Raquel, B. F., Ribeiro, M. A. O., and Nunes, F. L. S. (2025). “Fábrica de erros”: um arcabouço para simulação de erros na segmentação do ventrículo esquerdo em imagens cardíacas. In *Anais Estendidos do XXV Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS 2025)*, SBCAS Estendido 2025, page 25–30. Sociedade Brasileira de Computação (SBC).
- Raya, S. and Udupa, J. (1990). Shape-based interpolation of multidimensional objects. *IEEE Transactions on Medical Imaging*, 9(1):32–42.
- Ribeiro, M. A. O. (2022). *Segmentação do ventrículo esquerdo em exames de ressonância magnética cardíaca com aprendizado profundo e modelos deformáveis contendo restrições de forma*. PhD thesis, Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA).
- Ribeiro, M. A. O. and Nunes, F. L. S. (2021). Evaluating the pre-processing impact on the generalization of deep learning networks for left ventricle segmentation. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3505–3512.
- Ribeiro, M. A. O. and Nunes, F. L. S. (2022a). Abordagem híbrida adaptativa para segmentação do ventrículo esquerdo em exames de ressonância magnética cardíaca. Projeto de Pesquisa (Doutorado em Ciências) – Programa de Pós-graduação em Sistemas de Informação (PPgSI), Escola de Artes, Ciências e Humanidades da Universidade de São Paulo (EACH-USP).
- Ribeiro, M. A. O. and Nunes, F. L. S. (2022b). Left ventricle segmentation in cardiac MR: A systematic mapping of the past decade. *ACM Comput. Surv.*, 54(11s).

- Ribeiro, M. A. O. and Nunes, F. L. S. (2023). Left ventricle segmentation combining deep learning and deformable models with anatomical constraints. *Journal of Biomedical Informatics*, 142:104366.
- Robinson, R., Valindria, V. V., Bai, W., Oktay, O., Kainz, B., Suzuki, H., Sanghvi, M. M., Aung, N., Paiva, J. M., Zemrak, F., Fung, K., Lukaschuk, E., Lee, A. M., Carapella, V., Kim, Y. J., Piechnik, S. K., Neubauer, S., Petersen, S. E., Page, C., Matthews, P. M., Rueckert, D., and Glocker, B. (2019). Automated quality control in image segmentation: application to the UK biobank cardiovascular magnetic resonance imaging study. *Journal of Cardiovascular Magnetic Resonance*, 21(1):18.
- Robinson, R., Valindria, V. V., Bai, W., Suzuki, H., Matthews, P. M., Page, C., Rueckert, D., and Glocker, B. (2017). Automatic quality control of cardiac MRI segmentation in large-scale population imaging. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I*, page 720–727, Berlin, Heidelberg. Springer-Verlag.
- Sander, J., de Vos, B. D., and Išgum, I. (2020). Automatic segmentation with detection of local segmentation failures in cardiac MRI. *Scientific Reports*, 10(1).
- Shoab, M. A., Chuah, J. H., Ali, R., Hasikin, K., Khalil, A., Hum, Y. C., Tee, Y. K., Dhanalakshmi, S., and Lai, K. W. (2023). An overview of deep learning methods for left ventricle segmentation. *Computational Intelligence and Neuroscience*, 2023(1).
- Tavares, R. S., Sato, A. K., de Sales Guerra Tsuzuki, M., Gotoh, T., Kagei, S., and Iwasawa, T. (2010). Temporal segmentation of lung region mr image sequences using hough transform. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, page 4789–4792. IEEE.
- Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with maxwell's demons. *Medical Image Analysis*, 2(3):243–260.
- Valindria, V. V., Lavdas, I., Bai, W., Kamnitsas, K., Aboagye, E. O., Rockall, A. G., Rueckert, D., and Glocker, B. (2017). Reverse classification accuracy: Predicting segmentation performance in the absence of ground truth. *IEEE Transactions on Medical Imaging*, 36(8):1597–1606.
- Xue, W., Li, J., Hu, Z., Kerfoot, E., Clough, J., Oksuz, I., Xu, H., Grau, V., Guo, F., Ng, M., Li, X., Li, Q., Liu, L., Ma, J., Grinias, E., Tziritas, G., Yan, W., Atehortua, A., Garreau, M., Jang, Y., Debus, A., Ferrante, E., Yang, G., Hua, T., and Li, S. (2021). Left ventricle quantification challenge: A comprehensive comparison and evaluation of segmentation and regression for mid-ventricular short-axis cardiac MR data. *IEEE J. Biomed. Health Inform.*, 25(9):3541–3553.
- Yang, R., Mirmehdi, M., Xie, X., and Hall, D. (2013). Shape and appearance priors for level set-based left ventricle segmentation. *IET Computer Vision*, 7(3):170–183.
- Zou, K., Chen, Z., Yuan, X., Shen, X., Wang, M., and Fu, H. (2023). A review of uncertainty estimation and its application in medical imaging. *Meta-Radiology*, 1(1):100003.