Integrating Radiomics and Machine Learning to Improve Fluorescence Image Segmentation in in vitro models

Opening the Personal Gate between Technology and Health Care
M. Giacomini et al. (Eds.)

doi:10.3233/SHTI260125

Integrating Radiomics and Machine Learning to Improve Fluorescence Image Segmentation in in vitro models

Alessandra INTROVAIA[a,1], Gerardina RUOCCO[b], Letizia NICOLETTI[b], Samanta ROSATI[a], Valeria CHIONO[b] and Gabriella BALESTRA[a]
[a] *Department of Electronics and Telecommunications – Politecnico di Torino, Italy*
[b] *Department of Mechanical and Aerospace Engineering – Politecnico di Torino, Italy*
ORCiD ID: Alessandra Introvaia https://orcid.org/0009-0003-6526-7697
ORCiD ID: Samanta Rosati https://orcid.org/0000-0003-0620-594X
ORCiD ID: Valeria Chiono https://orcid.org/0000-0003-2067-7732
ORCiD ID: Gabriella Balestra https://orcid.org/0000-0003-2717-648X

Abstract. Myocardial infarction leads to fibrotic scar formation, compromising heart function and leading to heart failure. *In vitro* models of cardiac fibrotic tissue are essential tools for testing therapeutic strategies designed for this disease. Fluorescence microscopy is an imaging technique that provides high-resolution images of such models, but the segmentation of cellular components (e.g., cytoskeleton and nucleus) and therapeutic agents remains challenging. This early-stage study investigates the integration of radiomics and machine learning (ML) to improve cytoskeleton segmentation in fluorescence images of engineered cardiac fibrotic tissue. Experiments were conducted on a limited dataset of 18 fluorescence image triplets, and segmentation performance was evaluated qualitatively in the absence of manual ground truth annotations. Preliminary results show that the ML-based segmentation of the cytoskeleton can capture its complex morphological structures more effectively than traditional approaches.

Keywords. Image segmentation, Artificial Intelligence, Machine Learning, Fluorescence microscopy, *In vitro* models, Nanoparticles, Regenerative medicine.

1. Introduction

Cardiovascular diseases (CVDs) are a major global cause of mortality, with myocardial infarction (MI) being one of the most prevalent and severe conditions [1]. MI derives from a blockage of blood flow to the heart muscle, leading to the loss of several cardiomyocytes, followed by a cascade of inflammatory and reparative processes. Consequently, a fibrotic scar forms in the infarcted area, consisting of a stiff extracellular matrix (ECM) populated by cardiac fibroblasts, which preserves structural integrity, but it leads to a progressive heart failure.

In this context, predictive *in vitro* models of human pathological cardiac tissue are highly demanded for preclinical validation of advanced treatments, as they provide relevant testing platforms that mimic the biomimetic properties of post-infarct scar tissues.

[1] Corresponding Author: Alessandra Introvaia, alessandra.introvaia@polito.it.

Moreover, these in vitro models can be used to assess the efficacy of selective uptake and cardiac reprogramming by nanoparticle-mediated delivery of oligonucleotides [3].

Fluorescence microscopy (FM) enables high-resolution visualization of cellular structures, but automated analysis of FM images requires advanced image processing tools capable of extracting meaningful information from complex morphological patterns. Machine Learning (ML), a branch of Artificial Intelligence that enables algorithms to learn patterns directly from data without explicit programming [4], and Radiomics [5], which focuses on the extraction of quantitative features from biomedical images, provide complementary tools to address these challenges.

In this preliminary work, we explore how the integration of ML and radiomics enhances the characterization of cytoskeletal morphological organization, providing a more accurate and quantitative analysis of cellular structures.

2. Methods

2.1. Dataset

Engineered *in vitro* models of human cardiac fibrotic tissue were initially prepared using a recently published method [3]. Briefly, human adult ventricular cardiac fibroblasts (AHCFs) were cultured on electrospun nanofibrous scaffolds with biomimetic composition, stiffness and architecture. After 7 day-culture, they switched their phenotype into myofibroblasts with no need for additional biochemical stimulation by transforming growth factor b. Furthermore, the cells deposed an extracellular matrix rich in Collagen I and Fibronectin, in agreement with clinical data on post-infarct scar tissue. Non-functionalized and functionalized hybrid nanoparticles (i.e., NF-NPs and F-NPs) developed at Politecnico di Torino and PoliRNA Srl spinoff company were loaded with model siRNA-Cy5 (Cy5: Cyanine-5) to study cell uptake. The human cardiac tissue model samples were treated with NPs for 24 h and then, fixed with paraformaldehyde solution (4% v/v) in phosphate buffered saline (PBS) for 15 min, washed with PBS, and cells were permeabilized with Triton X-100 (0.5% v/v) in PBS for 8 min. Finally, samples were incubated for 30 min with Phalloidin and DAPI to stain cell cytoskeleton and nuclei, mounted on microscopy slides and observed under Nikon Ti2-E fluorescence microscope (Nikon Instruments) to detect siRNA-Cy5 signal. Cells were imaged at high magnification (60X).

Overall, the dataset employed in this study is made of 18 triplets of RGB uint8 images, 9 for the NF-NPs and 11 for the F-NPs treated cells. Each triplet consists of the 3 fluorescence channels: blue (DAPI, labelling nuclei), red (Cy5, marking siRNA-Cy5 loaded nanoparticles), and green (phalloidin–rhodamine, staining F-Actin in the cytoskeleton).

2.2. Segmentation pipeline

The first step of the analysis is to separate the three fluorescence channels to enhance segmentation precision and to better extract morphological information. Initially, the segmentation pipeline involved a heuristic and traditional image processing method, the Region Growing (RG) algorithm [6], that is based on intensity similarity criteria of neighboring pixels. RG was implemented in MATLAB with the seed threshold set to a

percentage of the maximum pixel intensity of the channel (50% for nuclei and 95% for cytoskeleton and siRNA channels), with neighborhood evaluation with a 50-pixel radius.

While RG produced reliable segmentations for nuclei and siRNA, it consistently under-segmented the cytoskeleton, due to the more complex morphology of these images, which highlighted the limitations of the traditional approach.

To overcome these limitations, the combination of ML and radiomics was introduced to improve the segmentation of cytoskeletal structures. Images were divided into a construction set (CS, 14 images) and a test set (TS, 4 images). The green channel of each image was normalized using min–max scaling and divided into small 5×5-pixel regions of interest (ROIs) labelled as background (class 0) or cytoskeleton (class 1) depending on the intensity of the pixels:

- If more than 90% of the pixels from a ROI showed an intensity higher than 0.75, the ROI was classified as cytoskeleton;
- If more than 90% of the pixels from a ROI showed an intensity lower than 0.05, the ROI was classified as background.

ROIs not satisfying either condition were excluded to ensure unambiguous class labels.

A total of 68432 and 24726 cytoskeleton ROIs and 1374142 and 387438 background ROIs were obtained for the CS and TS, respectively. From each ROI, 22 radiomic features [5] were extracted, including first-order statistical features and second-order texture features computed from the Gray Level Co-Occurrence Matrix (GLCM). All features were normalized using min-max scaling across the entire CS to ensure comparability between samples. Under-sampling of background ROIs (class 0) was performed to reduce class imbalance and ROIs in the CS were randomly split into a training set (TRS, 70% of ROIs, corresponding to 55276 class 0 and 44220 class 1 ROIs) and a validation set (VS, 30%, 23690 class 0 and 18951 class 1 ROIs) using a hold-out approach.

TRS and VS were used to construct a Random Forest (RF) classifier, with a fine-tuning process performed on key hyperparameters such as the number of trees, bagging ratio, and number of variables considered at each split. Finally, all ROIs in TRS, VS, and TS were classified, and the segmentation masks were constructed based on this classification. Since manual segmentation masks (ground truth) were not available, segmentation quality at the mask level was evaluated through qualitative visual assessment by comparing the results obtained from the RG and the ML-based approach applied to the green channel (cytoskeleton).

3. Results and Discussion

Figure 1 illustrates 3 results of RG applied to the red (siRNA-Cy5) and blue (nuclei) channels, achieving segmentation masks consistent with the corresponding fluorescence images, demonstrating its reliability for simpler morphological structures. However, RG applied to the green channel (cytoskeleton) yielded unsatisfactory results (Figure 2, first row): the segmentation masks appeared under-segmented compared to the fluorescence images, and characterized by fragmented and incomplete regions.

Figure 2 presents a comparison of segmentation masks for the same 3 samples showed in Figure 1, obtained using RG (first row) and the RF model (second row) (17 trees, bagging applied to 50% of the TRS, and random selection of attributes at each node

among 5 variables). Integrating the ML classifiers and radiomic features improves the quality of the green channel segmentation by detecting distant regions and segmenting less bright areas, which are typically caused by the nature of the fluorescent marker employed. Moreover, the ML method improved segmentation continuity and enabled the detection of spatially disconnected and low-intensity cytoskeletal regions. These qualitative improvements were consistently observed across multiple samples despite variability in signal intensity and morphology.

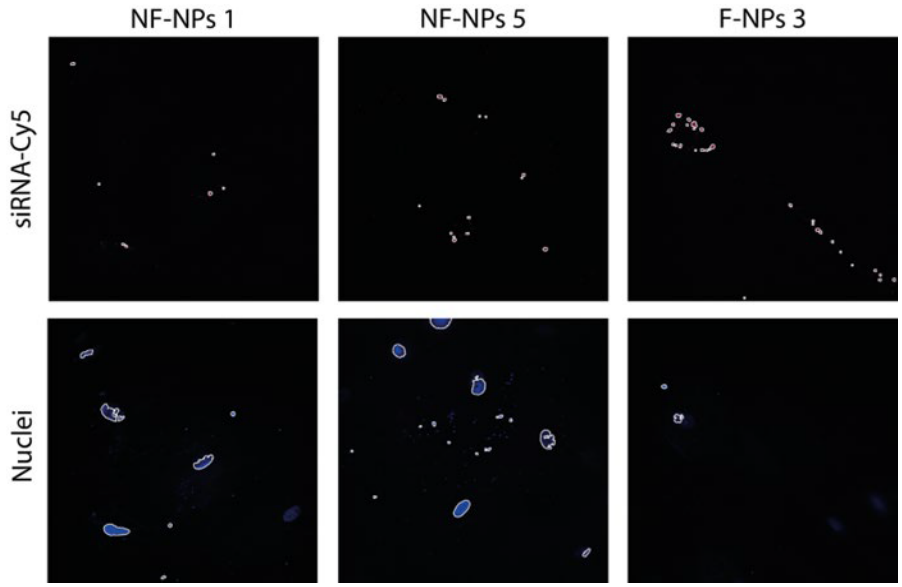

Figure 1. Segmentations of the blue and red channels with the RG algorithm on three image samples (NF-NPs 1, NF-NPs 5 and F-NPs 3). The first row refers to the siRNA-Cy5 component whereas the second one corresponds to the nuclei.

4. Conclusions

This paper presented the preliminary results of integrating radiomic features and ML into a segmentation pipeline developed for FM images of *in vitro* models of human cardiac fibrotic tissue. The proposed approach qualitatively demonstrated improved cytoskeleton segmentation compared to a traditional region growing method, particularly for low-intensity and spatially complex structures. More extensive studies are needed to optimize the proposed pipeline, for example by extending the image dataset to enhance the robustness and generalization ability, and by investigating the possible role of feature selection in improving system robustness. However, the proposed pipeline could be applied to quantitative studies of different nanoparticles internalization. This may involve extracting parameters such as the intensity distribution of siRNA-Cy5 pixels in relation to the nuclei and cytoskeleton, as well as the overall internalization rate. These metrics could provide in-sights into the efficiency and spatial distribution of nanoparticle uptake under different experimental conditions.

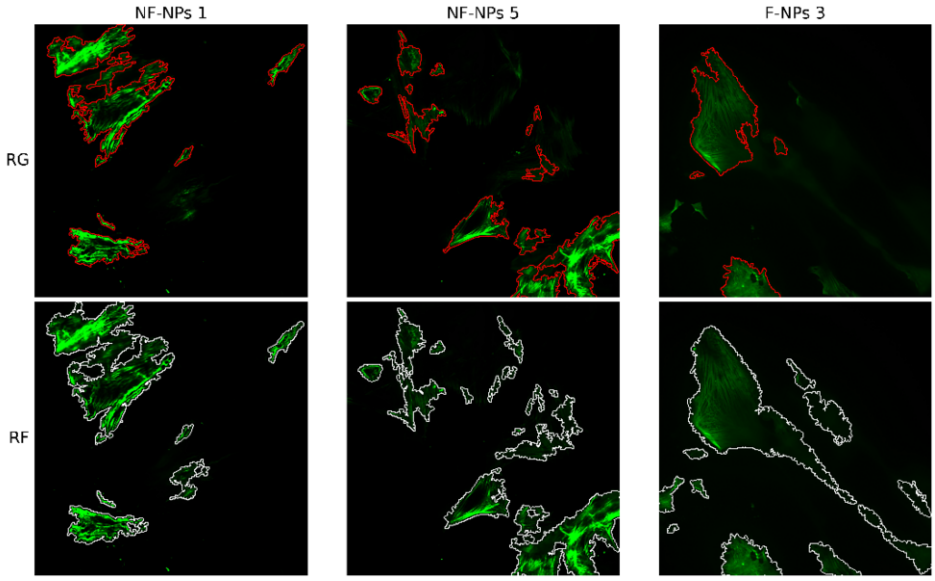

Figure 2. Comparison of the cytoskeleton segmentations (green channel) with RG algorithm (first row) and RF (second row) on three image samples (NF-NPs 1, NF-NPs 5 and F-NPs 3).

5. Acknowledgments

This study was funded by the European Union - Next Generation EU, Mission 4 - Component 2 through MERIT project (CN00000041_1; CUP E63C22000940007) and the NANOMIRC project (CN00000041; CUP C93C22002780006). This study was also carried out within the DESIRE project (CUP E53D23003950006) and the RECOVERY project (CUP E53D23017470001) - funded by European Union - Next Generation EU within the PRIN 2022 program (D.D. 104 - 02/02/2022 Ministero dell'Università e della Ricerca). NODES PoC project "NAPTER" supported by MUR – M4C2 1.5 of PNRR funded by the European Union - NextGenerationEU (ECS00000036).

References

[1] World Health Organization. Cardiovascular diseases. Geneva (CH): WHO; 2019. Available from: https://www.who.int/health-topics/cardiovascular-diseases

[2] Spedicati M, Tivano F, Zoso A, Bei J, Lavella M, Carmagnola I, et al. 3D bioartificial stretchable scaffolds mimicking the mechanical hallmarks of human cardiac fibrotic tissue. IJB. 2024 May 15;0(0):2247. doi: 10.36922/ijb.2247

[3] Ruocco G, Zoso A, Mortati L, Carmagnola I, Chiono V. Biomimetic Electrospun Scaffold-Based In Vitro Model Resembling the Hallmarks of Human Myocardial Fibrotic Tissue. ACS Biomater Sci Eng. 2023 Jul 10;9(7):4368–80. doi: 10.1021/acsbiomaterials.3c00483

[4] 1. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science. 2015 Jul 17;349(6245):255–60.

[5] Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. J Nucl Med. 2020 Apr;61(4):488–95. doi:10.2967/jnumed.118.222893

[6] Adams R, Bischof L. Seeded region growing. *IEEE Trans Pattern Anal Mach Intell.* 1994;16(6):641-647.