Exploring The Correlation and Relevance of Radiomics Features and Coronary Artery Calcification and Stenosis

Lida Alinezhad[1], Stefano Severi[1], Francesco Maffessanti[2], Cristiana Corsi[1]

[1]Department of Electrical Engineering and Information Engineering "Guglielmo Marconi", University of Bologna, Cesena, Italy
[2]Maria Cecilia Hospital, GVM Care and Research, Cotignola, Italy

Abstract

This study aims to assess the correlation and relevance of radiomic features extracted from coronary artery branches about the presence of calcification and stenosis. To do this, contrast-enhanced computed tomography (CECT) images from a subset of 54 cases were analyzed. Coronary arteries were segmented using the Chan-Vese segmentation method, followed by postprocessing in 3D Slicer. Radiomic features were then extracted using the PyRadiomics open-source library. Based on CECT reports, labels were assigned according to the presence of calcification and plaque. Feature selection was performed using correlation analysis and the Minimum Redundancy Maximum Relevance (MRMR) method. Finally, three different machine learning classification algorithms—Logistic Regression, Random Forest, and Support Vector Machines—were employed to evaluate the performance of the selected features. The results demonstrated that some features were informative for both calcification and stenosis. The features selected by correlation analysis combined with the Logistic Regression classifier showed more stable performance than the other methods.

1. Introduction

Coronary artery disease (CAD) remains a significant cause of morbidity and mortality worldwide [1]. Coronary computed tomography angiography (CCTA) has emerged as a non-invasive method for studying coronary atherosclerosis, with radiomics offering the potential for improved plaque characterization [2]. Radiomics, which involves the extraction of quantitative features from medical images, shows promise in predicting coronary artery calcification (CAC) and plaque risk. Studies have demonstrated that radiomics models that integrate CT-derived features with clinical risk factors can effectively predict calcified plaques [3]. Whole-heart radiomics obtained from low-dose CT has outperformed clinical variables in distinguishing Agatston scores and predicting coronary stenosis [4]. In the Framingham Heart Study, a radiomics-based score derived from CAC images enhanced risk prediction for major cardiovascular events, particularly in individuals with lower Agatston scores [5]. Furthermore, radiomics analysis of coronary plaques and perivascular adipose tissue have emerged as potential non-invasive biomarkers for cardiovascular risk stratification [6]. The combination of radiomic and clinical features in predictive models shows promise in supporting clinical decision-making processes for CAD assessment [7]. This advancement in radiomics applications may contribute to better understanding and management of coronary artery disease.

However, further research is required to address existing challenges and ensure the robustness and clinical translation of radiomics in the management of coronary artery disease. In this study, we aimed to evaluate the relationship between radiomics features and coronary artery calcification and plaque presence.

2. Methodology

2.1. Data Acquisition

The CECT dataset used in this study includes images of 54 subjects that were eligible for fine segmentation. The images are of size 512 × 512 pixels, with a pixel resolution of 0.32 × 0.32 mm. The slice spacing is 0.3 mm, and the slice thickness is 0.6 mm.

2.2. Image Segmentation

In this study, coronary arteries were segmented using the Chan-Vese segmentation algorithm [8]. Post-processing was then conducted in 3D Slicer to remove extraneous components and to separate the four branches of the coronary arteries: the right and left coronary arteries, the left anterior descending, and the circumflex artery. In general, from 514 segmented arteries, there were 49 positive calcification cases and 53 cases plaque severity more than 0.2 (>20% of stenotic vessel).

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.321

2.3. Feature Extraction

Using the open-source Python tool PyRadiomics, a set of quantitative radiomic features was extracted from previously segmented coronary arteries. In our application, these features were classified into two primary categories [9]: (1) Intensity, which encompasses first-order statistics derived from the intensity value histogram; (2) Texture, which assesses various textural properties such as heterogeneity, fineness, and coarseness, utilizing descriptors from established 3D texture metrics like the gray-level co-occurrence matrix (GLCM), run-length gray-level matrix (RLGL), and gray-level size-zone matrix (GLSZM).

2.4. Feature Selection and Classifier

In this study, two feature selection techniques were applied based on filter approaches: correlation analysis (CA) and Minimum Redundancy Maximum Relevance (MRMR) [10]. For classification, we utilized three classifiers: Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM). To investigate and compare different feature selection and classification methods, we created a three-dimensional parameter grid for the analysis. For both feature selection methods correlation and MRMR, we incrementally selected features ranging from 5 up to 15. These subsets of selected features were evaluated by using each of the three classifiers and area under ROC curves (AUC) [11].

The empirical stability of each classifier was quantified using the relative standard deviation (RSD) and a bootstrap approach. For each classification method, the model was trained on varying test sizes (0.2 and 0.3), with the training cohort being subsampled 100 times using the bootstrap method. RSD, expressed as a percentage, represents the absolute value of the coefficient of variation:

$$RSD\% = \frac{\sigma_{AUC}}{\mu_{AUC}}$$

where σ_{AUC} and μ_{AUC} are the standard deviation and mean of 100 AUC values respectively. It should be noted that higher stability in the case of classifiers corresponds to lower RSD values.

3. Results and Discussion

Regarding calcification prediction, in Figure 1, we show the results of the analysis for the three classifiers (LR in red, RF in green, and SVM in blue) using Correlation (top panels) and MRMR (bottom panels) feature selection methods at test sizes of 0.2, and 0.3.

SVM exhibited the highest RSD%, indicating lower stability, while LR and RF showed more consistent Performance. However, Logistic Regression consistently has the lowest RSD% across different numbers of features and test sizes. The RSD value decreases as the test size increases, suggesting that a larger test size may lead to more consistent model performance.

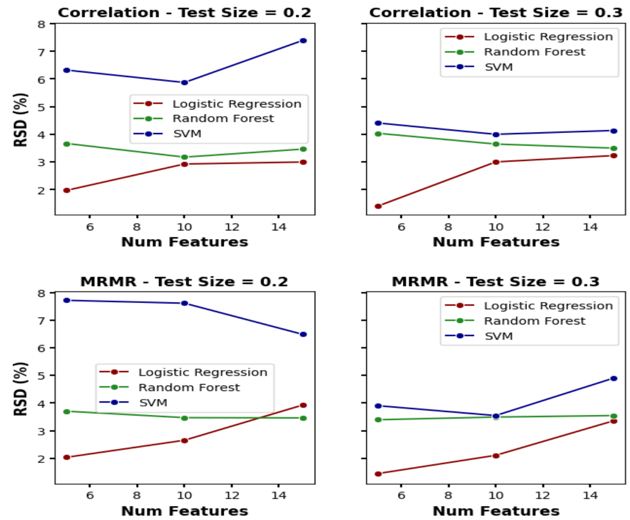

Figure 1. Comparison of (RSD%) for calcification prediction using three different classifiers across different numbers of features, employing Correlation (CA) and MRMR feature selection methods at varying test sizes (0.2 and 0.3).

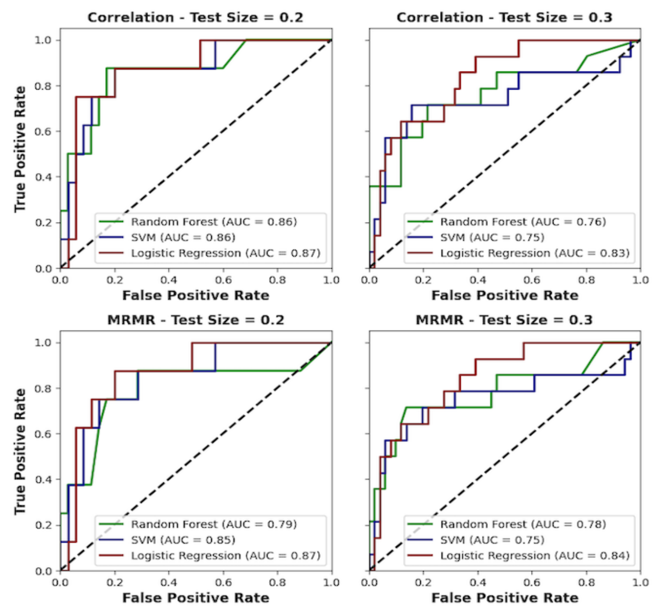

Figure 2. ROC curves comparing the performance of LR, RF, and SVM classifiers using Correlation and MRMR feature selection methods with 10 features for calcification study for calcification prediction.

Figure 2 presents ROC curves for LR, RF and SVM classifiers using CA and MRMR feature selection methods with 10 features in the calcification cohort, evaluated at different test sizes (0.2 and 0.3). The AUC values indicated that LR consistently outperformed the other classifiers across all conditions, particularly with higher AUCs using CA. SVM generally maintained moderate performance across all scenarios. Our results suggest that LR is more robust and effective in this context, while RF's performance was more sensitive to changes in test size.

The features that resulted to be more informative applying MRMR and CA were related to the intensity and texture are listed in table 1 for calcification:

Table 1. Features for calcification from CA and MRMR.

Calcification prediction
Firstoder_Skewness/GLCM_Correlation
Firstorder_Maximum/GLDM_dependenceVariance
NGTDM_Complexity/Firstorder_Skewness
Firstorder_Range/ NGTDM_Complexity
GLCM_Correlation/ Firstorder_Range

There are some features which are common between two selection method like: how pixel intensities are correlated (GLCM_Correlation), the asymmetry of the intensity distribution (Firstorder_Skewness), the complexity of texture variations (NGTDM_Complexity), and the overall intensity contrast (Firstorder_Range).

Regarding stenosis prediction, we performed a similar analysis.

Figure 3 displays the RSD for LR (in red), RF (in green), and SVM (in blue) classifiers across different numbers of features using CA (top panels) and MRMR (bottom panels) feature selection methods, evaluated at test sizes of 0.2, and 0.3.

Like calcification, SVM showed the highest RSD%, indicating lower stability for stenosis prediction, particularly when using the CA method for feature selection. RF exhibited the most consistent stability across both feature selection methods with 10 features. For stenosis prediction, the most stable classifier varied depending on the test size and number of features. These results underscore the importance of selecting an appropriate feature selection method and the optimal number of features to ensure stable model performance.

Figure 4 shows ROC curves for LR, RF, and SVM classifiers using CA and MRMR feature selection methods with 10 features at different test sizes (0.2 and 0.3). LR consistently achieved the highest AUC values with CA method.

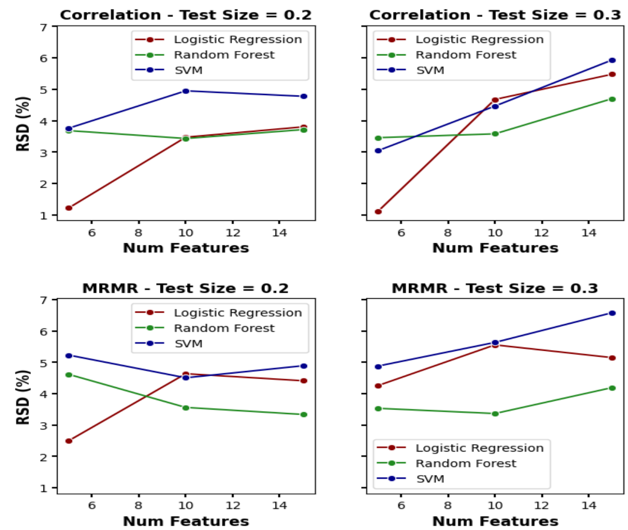

Figure 3. RSD (%) for LR (in red), RF (in green), and SVM (in blue) classifiers across different numbers of features, using CA and MRMR feature selection methods at varying test sizes (0.2 and 0.3) for stenosis prediction.

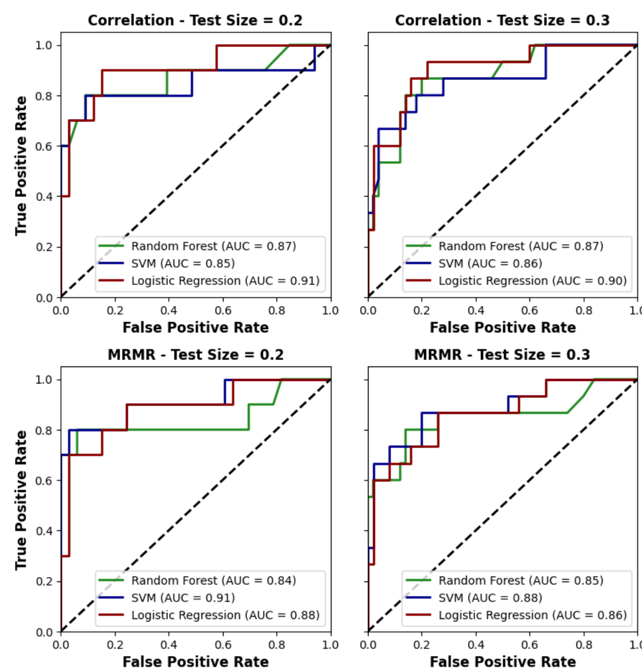

Figure 4: ROC curves comparing the performance of Logistic Regression, Random Forest, and SVM classifiers using Correlation and MRMR feature selection methods with 10 features for stenosis prediction.

The informative features selected by applying CA and MRMR are listed in Table 2. There are some similarities between the features in CA and MRMR: (GLCM_ClusterShade) that the skewness and nonuniformity in texture patterns, (NGTDM_Complexity) the variability in intensity differences between neighboring pixels, and (Firstorder_Skewness), the asymmetry of the intensity distribution.

Table 2 features for Stenosis from CA and MRMR.

Stenosis Prediction
Firstorder_Maximum / Firstorder_Skewness
Firstorder_Range /Firstorder_Uniformity
GLCM_ClusterShade/ GLCM_ClusterShade
NGTDM_Complexity / NGTDM_Complexity
Firstorder_Skewness / NGTDM_Strength

Both predictions share the asymmetry of the intensity distribution and the variability in intensity differences between neighboring pixels in the top five features. Overall, our results underscore the importance of choosing an appropriate classifier and feature selection method to optimize model performance.

4. Conclusion

In conclusion, Logistic Regression with Correlation Analysis appears to be the most effective combination of predicting calcification and stenosis, providing the highest stability and classification performance across different test sizes, with AUC values reaching up to 0.91 for both prediction scenarios. The classifier and feature selection methods are crucial in enhancing the prediction accuracy of CAD patients. Furthermore, there are some features which are repeated for classification and stenosis with both selection methods which indicates that these features could have the potential for coronary artery calcification and plaque formation predictions.

However, further testing using larger and more diverse datasets is essential to validate these results. To further enhance the robustness and interpretability of the model, the use of explainable AI techniques could be explored, helping clinicians to better understand the decision-making process of predictive models.

Acknowledgements

This work was supported by the Italian National Recovery and Resilience Plan (PNRR). We gratefully acknowledge funding from the European Union – NextGenerationEU, under the framework of the PNRR, managed by the Italian Ministry of Economy and Finance.

References

[1] Murgia, A., et al. “Cardiac Computed Tomography Radiomics: An Emerging Tool for The Non-invasive Assessment of Coronary Atherosclerosis.” Cardiovascular Diagnosis and Therapy, 10(6), 2005-2017, 2020.

[2] Kim, Justin N., et al. “Pericoronary Adipose Tissue Radiomics from Coronary Computed Tomography Angiography Identifies Vulnerable Plaques.” Bioengineering 10.3 (2023).

[3] Hu, Guo-qing, et al. “Predicting Coronary Artery Calcified Plaques Using Perivascular Fat CT Radiomics Features and Clinical Risk Factors.” BMC Medical Imaging 22.1 (2022).

[4] Homayounieh, Fatemeh, et al. “Prediction of Coronary Calcification and Stenosis: Role of Radiomics from Low-Dose CT”. Academic radiology 28.7 (2021).

[5] Eslami, Parastou, et al. "Radiomics of Coronary Artery Calcium in the Framingham Heart Study." Radiology: Cardiothoracic Imaging 2.1 (2020).

[6] Corti, Anna, et al. "Enhancing Cardiovascular Risk Stratification: Radiomics of Coronary Plaque and Perivascular Adipose Tissue Current Insights and Future Perspectives.” Trends in Cardiovascular Medicine (2024).

[7] Militello, Carmelo, et al. "CT Radiomic Features and Clinical Biomarkers for Predicting coronary artery disease." Cognitive Computation 15.1 (2023).

[8] Chan, Tony, and Luminita Vese. "An Active Contour Model without Edges." International Conference on scale-space theories in computer vision. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999.

[9] Coroller, Thibaud P., et al. "CT-based Radiomic Signature Predicts Distant Metastasis in Lung Adenocarcinoma”. Radiotherapy and Oncology 114.3 (2015).

[10] Brown, Gavin, et al. "Conditional Likelihood Maximization: A Unifying Framework for Information Theoretic Feature Selection." The journal of machine learning research 13.1 (2012).

[11] Parmar, Chintan, et al. "Machine Learning Methods for Quantitative Radiomic Biomarkers." Scientific reports 5.1 (2015).

Address for correspondence:
Lida Alinezhad
University of Bologna, Via dell’Università 50,
47522 Cesena (FC), Italy.
lida.alinezhad2@unibo.it