ISSN: 1692-7257 - Volume 1 – Number 47 - 2026

RCTA
Revista Colombiana de Tecnologías de Avanzada
UNIPAMPLONA

Digital Object Identifier: 10.24054/rcta.v1i47.4285

Comparison of artificial intelligence models for heart disease detection using magnetic resonance imaging

Comparación de modelos de inteligencia artificial para la detección de cardiopatías usando imágenes de resonancia magnética

PhD. Jorge Eliecer Gómez Gómez[1], Ing. Juan José Gúzman Pineda[1]
MSc. Velssy Liliana Hernández Riaño[1]

[1] ***Universidad de Córdoba***, *Departamento de Ingeniería de sistemas, grupo Sócrates, Montería, Colombia.*

Correspondence: *jeliecergomez@correo.unicordoba.edu.co*

Received: *July 21, 2025.* ***Accepted:*** *December 20, 2025.* ***Published:*** *January 01, 2026.*

How to cite: *J. E. Gómez Gómez, J. J. Guzmán Pineda y V. L. Hernández Riaño, "Comparación de modelos de inteligencia artificial para la detección de cardiopatías usando imágenes de resonancia magnética", RCTA, vol. 1, n.º. 47, pp. 34-45, Jan. 2026. Recuperado de https://ojs.unipamplona.edu.co/index.php/rcta/article/view/4285*

Abstract: In this study, three artificial intelligence techniques were compared for the binary detection (healthy/diseased) of cardiopathies using axial slices from cardiac magnetic resonance imaging. Using a dataset of 150 patients in NIfTI format, the images were preprocessed (normalization, rescaling to 128 × 128, RGB conversion, and data augmentation) and split on a per-patient basis using an 80/20 ratio. A Random Forest model with GLCM radiomic descriptors and first-order statistical features, a convolutional neural network (CNN), and a YOLOv8-based model adapted for binary classification were evaluated. The models were compared using accuracy, precision, recall, F1-score, and AUC, and explainability techniques (SHAP, Grad-CAM, Integrated Gradients, and occlusion sensitivity) were applied to validate the anatomical coherence of the predictions. Overall, the results indicate that pretrained deep learning approaches, such as YOLOv8, offer substantial advantages in terms of accuracy and interpretability, positioning them as a promising alternative for the development of intelligent decision-support systems for the diagnosis of structural cardiopathies.

Keywords: heart disease, MRI, YOLOv8, Random Forest, CNN, artificial intelligence.

Resumen: En este trabajo se compararon tres técnicas de inteligencia artificial para la detección binaria (sano/enfermo) de cardiopatías mediante cortes axiales de resonancia magnética cardíaca. Usando un conjunto de 150 pacientes en formato NIfTI, las imágenes se preprocesaron (normalización, reescalado a 128 × 128, conversión a RGB y aumento de datos) y se dividieron en una proporción 80/20 por paciente. Se evaluaron Random Forest con descriptores radiómicos GLCM y estadísticos de primer orden, una red neuronal convolucional (CNN) y un modelo basado en YOLOv8 adaptado a clasificación binaria. Los modelos se compararon mediante accuracy, precision, recall, F1-score y AUC, y se aplicaron técnicas de explicabilidad (SHAP, Grad-CAM, Integrated Gradients y sensibilidad a la oclusión) para validar la coherencia anatómica de las predicciones. En conjunto, los resultados indican que los enfoques de aprendizaje profundo preentrenado, como YOLOv8, ofrecen ventajas sustanciales en términos de precisión e interpretabilidad,

posicionándose como una alternativa prometedora para el desarrollo de sistemas inteligentes de apoyo al diagnóstico de cardiopatías estructurales.

Palabras clave: cardiopatías, resonancia magnética cardíaca, YOLOv8, Random Forest, CNN, inteligencia artificial.

1. INTRODUCTION

Cardiovascular diseases remain the leading cause of death worldwide, accounting for approximately 32% of all global deaths each year, according to the World Health Organization (WHO) [1]. It is estimated that more than 17.9 million people die annually from heart and blood vessel-related causes, a figure that continues to rise owing to population aging, changes in lifestyle, and limited coverage of preventive strategies. The WHO has emphasized the need for early interventions, especially in low- and middle-income countries, where the burden of cardiovascular disease has shown an alarming upward trend.

This situation is no different in Colombia. The Ministry of Health and Social Protection has reported that in recent years, diseases of the circulatory system have consistently ranked among the leading causes of death in the country, with ischemic heart disease, stroke, and hypertensive diseases being the most prevalent [2]. In 2022, the cardiovascular disease mortality rate was estimated at approximately 176 deaths per 100,000 inhabitants, with a higher burden among older adults and in rural areas [3].

Despite advances in hospital infrastructure and coverage, structural challenges related to equitable access to timely diagnosis persist, particularly in regions with limited specialized coverage. In this context, the adoption of technologies capable of facilitating the early and efficient detection of cardiovascular diseases, such as artificial intelligence-based solutions, has emerged as a public health priority [4].

In this scenario, the early and accurate detection of cardiac conditions, even before they manifest clinically, has become a key objective of modern medicine. Among the available diagnostic tools, cardiac magnetic resonance imaging (MRI) stands out for its ability to provide detailed images of the heart's morphology and function without exposure to ionizing radiation [5], [6]. However, the traditional analysis of these images relies on visual interpretation by experts, a practice subject to inter-observer variability, is time-consuming, and is not always accessible in all clinical settings [7].

In this context, artificial intelligence (AI) has emerged as a promising solution for automating diagnostic processes, reducing clinical workloads, and improving analysis consistency [8]. Multiple deep learning-based models have been developed that allow for the segmentation of cardiac chambers, quantification of volumes, and classification of specific pathologies from medical images [9], [10]. However, in daily clinical practice, many initial decisions are based on the fundamental question: Is there a cardiac condition that warrants attention?

Addressing this question from a binary perspective—presence or absence of disease—is especially relevant in scenarios such as population screening, emergency department triage, and telemedicine [11], [12]. In such contexts, lightweight, interpretable, and efficient models capable of providing reliable answers from basic images are crucial for reducing diagnostic times and appropriately prioritizing available clinical resources [13].

Beyond technical performance, the reliability of these systems is fundamental to their adoption. Several studies have shown that AI models can achieve levels of accuracy comparable to or exceeding those of human experts, particularly when trained on well-labeled and heterogeneous datasets. However, their implementation requires not only quantitative validation but also mechanisms that allow for the explanation and justification of each clinical decision to build trust among healthcare professionals and patients.

The incorporation of explainability techniques has proven to be an effective strategy for improving the transparency of these systems, making it possible to identify the regions of the image that most influence the prediction [12]. This not only facilitates medical audits and second opinions but also supports the ethical and legal integration of AI into structured healthcare systems, particularly in processes such as automated screening, computer-assisted diagnosis, and remote second readings.

This study focused on the development and validation of MRI-based binary detection models for heart disease using deep learning. Through the implementation of machine learning and neural networks, we aim to offer an effective, reproducible, and clinically viable mechanism to support early decision-making in cardiology, with a particular focus on high-demand or resource-limited settings.

2. RELATED WORKS

This section critically analyzes the most relevant studies grouped into five thematic lines that represent the main current research trends.

2.1. Automatic Segmentation of Cardiac Structures

One of the most established lines of research in the literature is the automatic segmentation of cardiac structures from magnetic resonance imaging (MRI). In this field, the work of [7] marked a milestone by presenting the results of the Automatic Cardiac Diagnosis Challenge (ACDC), where various deep learning methods applied to the segmentation of the left ventricle (LV), right ventricle (RV), and myocardium (MYO) were assessed.

This benchmark established a standard in the area, achieving average DICE coefficients of 0.95 for the LV, 0.89 for the RV, and 0.89 for the MYO, primarily using U-Net-based architectures. Despite these advances, the authors pointed out limitations regarding the generalizability of the models outside the ACDC set, which prompted subsequent research focused on robustness and transferability.

In response to these limitations, [14] proposed a convolutional network that integrates directional feature maps (DFMs) designed to capture oriented spatial patterns characteristic of cardiac structures, such as curved edges and boundaries between the myocardium and ventricular chambers.

This approach overcame the limitations of isotropic architectures, such as U-Net, achieving Dice coefficients of 0.955, 0.901, and 0.888 for LV, RV, and MYO, respectively, on the ACDC set. The added value of this architecture lies in its ability to address both interclass indistinction and intraclass variability, which are critical aspects for reliable clinical diagnosis.

Meanwhile, [15] developed Heart-Net, a multimodal architecture that combines MRI images with physiological signals (ECG and blood pressure), integrating a CNN and LSTM in a late fusion approach.

This approach not only improves the accuracy of cardiac structure segmentation but also allows for the classification of cardiac pathologies with an accuracy exceeding 98%. The combination of morphological and functional information simulates medical diagnostic reasoning and has proven to be highly robust to anatomical variations and clinical noise.

Complementing these proposals, [16] introduced a hybrid computer-aided diagnosis (CAD) system that employs ResU-Net to segment cardiac structures and a Vision Transformer (ViT) to classify infarcted regions. This approach was particularly effective in identifying damaged myocardial tissue, achieving an AUC of 0.98 and an overall diagnostic accuracy of 97.3%.

The ability of transformers to capture long-range spatial relationships makes them ideal for detecting diffuse and subtle patterns, which is fundamental for the early diagnosis of myocardial infarction.

2.2. Automatic Classification of Cardiac Diseases

The direct classification of cardiac pathologies from MRI represents a significant evolution compared to traditional approaches that rely on multiple intermediate stages. In [11], SA-YOLO, an adapted version of YOLOv8 with spatial attention modules was presented.

This model allows for the detection of pathologies directly from images without the need for prior segmentation. It achieved an overall accuracy of 98.1% and an inference time of 23 ms per image, making it viable for real-time use in clinical settings with limited computational resources. Meanwhile, in [17], an innovative approach was introduced using strain fields between the diastolic and systolic phases to encode the contractile dynamics of the heart. Using an architecture based on autoencoders and an SVM classifier, they achieved an average accuracy of 94.5.

This type of functional input allows the identification of contractile dysfunctions characteristic of diseases such as dilated cardiomyopathy (DCM) and hypertrophic cardiomyopathy (HCM), providing a more informative alternative to static images.

In another noteworthy study, [18] applied Vision Transformers in combination with explainability techniques, such as Grad-CAM, to detect myocarditis. Their model achieved an F1-score of 0.9741 and an AUC of 0.9616, which was notable for both its accuracy and interpretability.

This study represents a significant advance by explicitly integrating the visualization of active regions that motivate classification, thereby improving the transparency of the automated diagnostic process.

Finally, [12] adopted a complementary approach by integrating algorithms such as XGBoost and LightGBM on multi-source clinical variables without relying exclusively on MRI. Although they did not segment images, they achieved an average AUC of 0.989, demonstrating that combining structured data with high-performance models can offer reliable predictions, even in the absence of detailed imaging.

2.3. Explainability and Interpretability of Models

The acceptability of artificial intelligence models in clinical contexts depends not only on their accuracy but also on their interpretability. In this regard, [19] proposed a co-learning architecture that simultaneously segments the myocardium and detects post-infarction scar tissues.

The model incorporates an attentional mechanism that highlights relevant regions in the predictions, achieving a sensitivity of 91.8% and diagnostic accuracy of 94.6%. The spatial coherence between the segmentation and functional detection tasks improves the clinical confidence in the results.

Likewise, the work of [18] stands out for incorporating Grad-CAM as an explainability tool in a sensitive context, such as the diagnosis of myocarditis. These techniques allow for the visual and quantitative validation of the model's decisions, showing which specific regions influenced the prediction, thereby strengthening transparency and medical confidence.

2.4. Clinical Validation and Adoption in Hospital Settings

One of the barriers to the adoption of AI models in medicine is the lack of clinical validation in real-world settings. In this regard, [5] conducted a prospective study of 61 patients comparing conventional cine sequences with those reconstructed using deep learning.

Their model demonstrated excellent agreement with traditional functional measures ($r > 0.98$) and reduced the acquisition time by 50%, improving visual quality, especially in subendocardial structures.

On the other hand, [20] and [6] conducted systematic reviews of the use of AI in cardiovascular imaging, addressing topics such as integration into hospital workflows, regulation by entities such as the FDA, and model traceability. These reviews highlight that, beyond technical performance, the clinical acceptance of a model depends on its transparency, acceptable inference time, and compatibility with HIS/PACS.

2.5. Preventive Diagnosis and Multimodal Approaches

The application of AI in medicine is not limited to current diagnosis but extends to the preventive prediction of cardiovascular risks. In [21], researchers demonstrated that deep neural networks can predict future cardiac events from liver images with an AUC of 0.87 by combining abdominal MRI and clinical variables using a multimodal transformer. This approach breaks the traditional logic of structural diagnosis and allows for the exploration of latent correlations in other imaging modalities.

Similarly, in [22], researchers trained a ResNet-type CNN with care modules to classify five cardiac pathologies directly from cine MRI without prior segmentation. Their model achieved 96.4% accuracy and an inference time of less than one second, making it a viable option for clinics with high workloads and limited specialized staff.

Finally, in [13], the researchers synthesized advances in AI for myocardial diseases, highlighting the emerging role of generative models, recurrent neural networks, and virtual simulations. By integrating multiple temporal and morphological dimensions, these approaches broaden the diagnostic spectrum and suggest new avenues for personalizing treatment and simulating clinical progression.

3. MATERIALS AND METHODS

3.1. Dataset Description

The dataset used consisted of cardiac magnetic resonance imaging studies stored in NIfTI format, obtained from [7], with 150 patients divided into five subgroups (30 normal subjects, NOR; 30 patients with previous myocardial infarction; 30 patients with dilated cardiomyopathy; 30 patients with hypertrophic cardiomyopathy; 30 patients with abnormal right ventricle). Each volume includes axial slices from different points in the cardiac cycle (e.g., frame01 and frame12) accompanied by manually segmented masks (_gt.nii.gz). These masks delineate structures such as the myocardium, left ventricle (LV), and right ventricle (RV).

Min-max normalization was applied for each patient (based on the intensity values of each volume). Axial slices were interpolated to 128×128 or 256×256 pixels using bicubic interpolation, background noise was removed with adaptive thresholding, and the intensities were converted to 8 bits. For YOLOv8, the images were duplicated into three RGB channels for compatibility. The same pipeline was uniformly applied to the training and validation datasets.

The images were obtained from a clinical repository structured in subfolders per patient, with anatomical and temporal variability. This richness allows for the training of robust models for binary classification (presence or absence of disease) and experimentation with segmentation tasks. The dataset was assembled from Google Drive in a collaborative environment using Google Colab software.

3.2. General Processing Flow

As a preliminary step in the implementation of each model, a generic processing flow was defined to ensure data quality and consistency. This workflow, represented in Fig. 1, begins with a cleaning and normalization process of the cardiac magnetic resonance imaging volumes, followed by a preprocessing module that includes cropping the region of interest, selecting representative axial slices, rescaling to standard resolutions (128×128 or 256×256 pixels), and converting to formats compatible with each architecture (grayscale or RGB).

Subsequently, feature selection and transformation techniques were applied according to the approach used (flattening in Random Forest, convolutions in CNN, and direct RGB processing in YOLOv8). Finally, each model was trained and evaluated independently using the same performance metrics to allow objective comparison. This structured approach ensures pipeline traceability and experimental replicability.

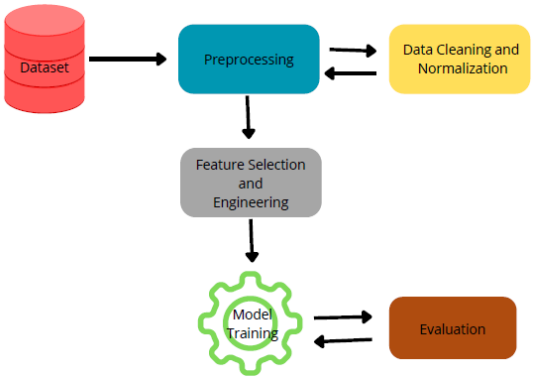

Fig. 1. *General processing flow*

3.3. Data Preparation

The dataset used consists of three-dimensional volumes .nii.gz format, which were divided 80/20 at the patient level, grouping all slices from each subject into the same set with stratification by class (healthy/disease). This prevented information leakage and ensured independence between the training and validation. The data were preprocessed to obtain representative 2D axial slices for each of the studies.

The preprocessing steps included the following:

- Extraction of the mid-axial slice from the volume, assuming that the plane provides a representative view of the heart.

- Normalization of pixel intensities to the range [0, 1] using min-max scaling, with the aim of reducing variability arising from heterogeneous acquisition conditions.

- Uniform resizing of all images to a fixed size of 128×128 pixels.

- Conversion to flat format: Because the random forest model requires vectorized inputs, each image was converted from its matrix form (128,128,1) to a one-dimensional vector of 16,384 features using flattening.

This transformation allowed each image to be represented as a point in a high-dimensional feature space, accompanied by a binary label corresponding to the patient's clinical condition (healthy = 0, sick = 1).

To improve the discriminative capacity, radiomic features based on GLCM texture (contrast, dissimilarity, homogeneity, energy, and correlation) and first-order statistics (mean, variance, kurtosis, and skewness) were extracted. These descriptors were calculated per slice using the scikit-image library and served as inputs to the Random Forest.

3.4. Data Augmentation

Given the small size of the dataset and its slightly unbalanced distribution, a data augmentation strategy was applied using the ImageDataGenerator. This technique allows for increased diversity in the training set through random transformations without the need to collect new images. The transformations included random rotations (±10°), random zoom (±10%), horizontal and vertical shifts (±10%), and horizontal flipping. These operations preserved the relevant anatomical features and improved the generalizability of the model.

4. RESULTS

4.1. Random Forest Model Performance

This model was trained and evaluated using flattened images from the axial slices of cardiac magnetic resonance imaging. After training, its performance was evaluated on a previously separated validation set (20% of the total samples), and the following quantitative results were obtained:

Overall Accuracy: 0.8000
Precision: 0.8000
Recall: 1.0000
F1-score: 0.8889
Area under the ROC curve (AUC): 0.6424

Figure 2 shows the ROC curve of the model. The area under the curve (AUC = 0.6424) indicates a moderate discriminatory ability of the model to distinguish between healthy and diseased patients. Although the sensitivity is high, the presence of false positives reduces the classifier's overall accuracy.

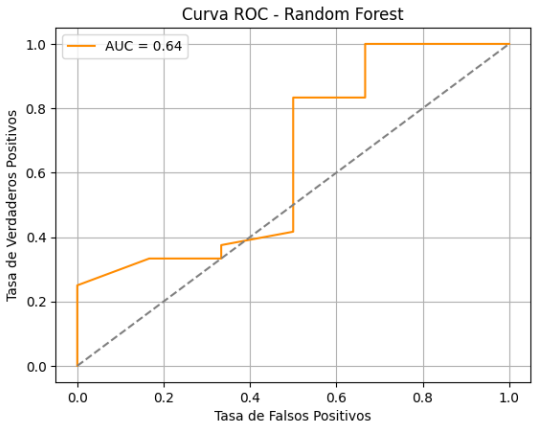

Fig. 2. *ROC curve of the Random Forest model.*

The Random Forest model's performance is characterized by a high capacity for detecting positive (ill) cases, making it a potentially useful tool in contexts where the omission of critical cases is unacceptable (high sensitivity). However, the number of false positives recorded suggests that it may generate unnecessary alerts in some healthy patients, which should be considered in real-world clinical settings. This behavior can also be attributed to the loss of spatial information during the image flattening process, which limits the model's ability to differentiate subtle anatomical patterns.

4.2. CNN Model Performance

The model was trained and evaluated using a dataset of 30 samples. The results obtained by the convolutional neural network are as follows:

Overall Accuracy: 0.7000
CHD Class Accuracy: 0.8000
CHD Class Recall: 0.8333
CHD Class F1 Score: 0.8163

Figure 3 shows the confusion matrix of the CNN model. The neural network correctly identified 20 cases with CHD and one healthy case (non-CHD). However, it produced five false positives, incorrectly classifying healthy patients as having CHD, and four false negatives, corresponding to patients with CHD who were classified as healthy. This behavior reflects a slight bias in the model towards the positive class, prioritizing the detection of pathological cases over the correct identification of healthy participants.

This trend may be attributed to an imbalance in the dataset (a greater number of images of sick patients) or the limited capacity of the model to generalize

and learn representative patterns of normal myocardium.

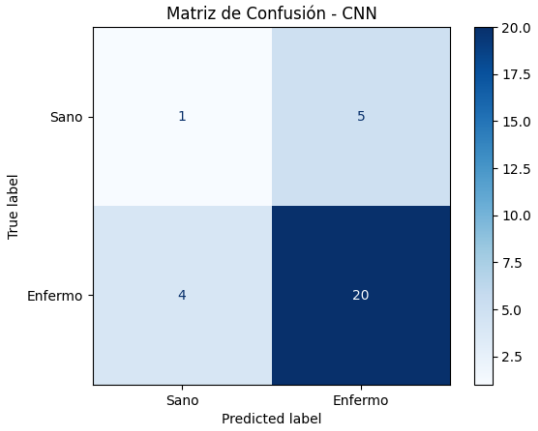

Fig. 3. *Confusion matrix of the CNN model.*

The model's ROC curve, which shows an AUC of 0.4167, indicates discriminatory power below the limit of chance, which is not acceptable in the healthcare field. Although the network achieved acceptable sensitivity, its diagnostic utility was limited by the low AUC and inability to correctly classify negative cases.

Despite its high sensitivity for detecting CHD cases, the CNN model exhibited significant deficiencies in identifying healthy patients, which could result in a high number of false positives in a clinical setting. This behavior may be related to the imbalance of the dataset and the possible need for further adjustments to the architecture or regularization methods. Future versions of the model should incorporate techniques such as focal loss, class weighting, or specific enhancements for minority classes.

4.3. Performance of the YOLOv8

Model The performance achieved by this model was as follows.

Overall Accuracy: 0.8000
Precision: 0.8000
Recall: 1.000
F1-score: 0.8900

Figure 4 shows the ROC curve obtained from the classification probabilities of the YOLOv8 model for the positive (ill) class. The area under the curve (AUC) was 0.83, reflecting a good, though not outstanding, discrimination ability. The curve has an upward slope and partially approaches the upper-left vertex, indicating that the model achieves a high true positive rate with a moderate number of false positives. This result confirms that YOLOv8 can effectively distinguish between healthy and ill subjects, although its performance could be improved through more precise calibration of the decision threshold or a better balance between classes.

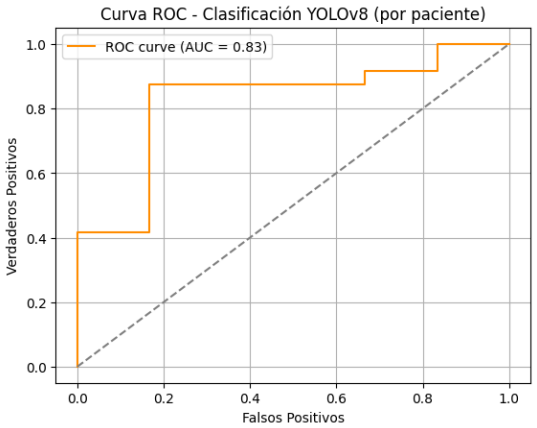

Fig. 4. *ROC curve of the YOLOv8 model.*

The model showed superior performance compared to the other approaches evaluated in terms of accuracy and class balance. Its ability to generalize correctly across a diverse validation set, coupled with its low error rate and high AUC score, positions it as a highly viable alternative for the automated detection of congenital heart disease in medical images. The model's stability also suggests effective learning of discriminative patterns without overfitting, even in a clinical setting with moderate resource availability.

4.4. Model Explanability

4.4.1. Random Forest

To understand how the model makes its decisions and validate the physiological relevance of the features used, two complementary approaches to global and local explainability were implemented. SHAP (Shapley Additive Explanations), based on cooperative game theory, and Permutation Importance, which evaluates the variation in model performance when each variable is randomly altered. Both methods were applied to the validation set using the six radiomic descriptors obtained from the gray-level co-occurrence matrix (GLCM): Homogeneity, Energy, ASM (Angular Second Moment), Contrast, Dissimilarity, and Correlation.

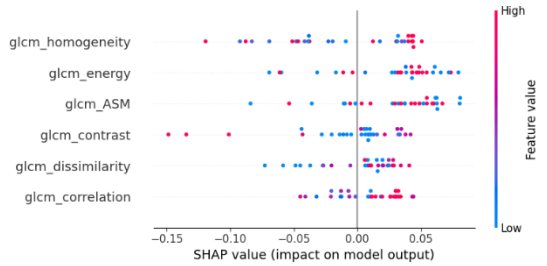

Fig. 5. *Impact of the model on the outputs*

Analysis of the mean absolute SHAP values (see Fig. 5) identified the following variables as having the greatest impact on prediction:

• GLCM_Homogeneity (0.0449)
• GLCM_Energy (0.0442),
• GLCM_ASM (0.0436).

These three characteristics are associated with the uniformity of the myocardial tissue. Patients with high homogeneity and energy values tended to be classified as healthy, whereas more heterogeneous textures (low homogeneity and high contrast) were linked to the diseased class. The variables Contrast, Dissimilarity, and Correlation had a smaller contribution, although their combined effect helped distinguish mild structural abnormalities. The representation of individual SHAP values showed a consistent trend: negative homogeneity values decreased the probability of being healthy, and positive values increased it, confirming that the model internalized a clinically consistent pattern.

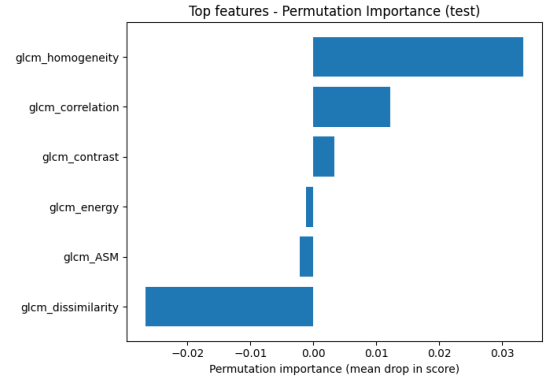

Fig. 6. *Importance of permutation.*

Regarding the importance of permutation, Figure 6 shows the average importance and standard deviation obtained using the permutation method. The results partially coincided with SHAP; homogeneity remained the most relevant (0.0333), followed by correlation (0.0122) and contrast (0.0033). However, variables such as ASM and Energy showed slight negative variations, indicating some redundancy between them, which is expected, given that both reflect uniformity and smooth texture. The agreement between SHAP and Permutation suggests a high internal consistency in the model. The dominance of uniformity descriptors implies that the model associates intensity dispersion with pathology, consistent with the typical structural changes in cardiomyopathy (hypertrophy, fibrosis, or remodeling). Finally, because Random Forest does not operate directly on the image but on global feature vectors, its spatial interpretability is limited; however, it offers a clear conceptual explanation based on the statistical properties of the tissue.

4.4.2. CNN

The convolutional neural network (CNN) model was analyzed using three complementary visual explainability techniques:

- Grad-CAM (Gradient-weighted Class Activation Mapping)
- Saliency Maps
- Integrated Gradients (IG).

These tools allow the identification of image regions that most influence the final prediction, providing a direct spatial interpretation of the model's attention. The Grad-CAM technique (see Fig. 7) generated heat maps indicating which regions of the axial slice most intensely activated the convolutional layers when classifying a patient as having a disease.

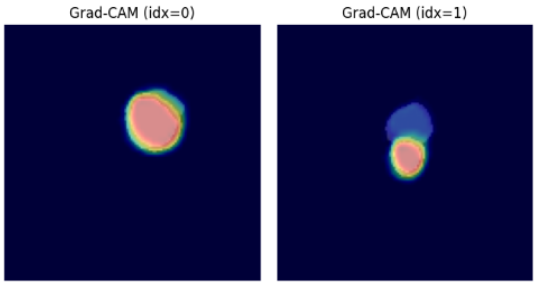

Fig. 7. *Grad-CAM heat maps*

In most positive cases, the red and yellow areas were concentrated in the left ventricle and interventricular septum, which are commonly affected in hypertrophic and dilated cardiomyopathies. In healthy patients, the activations were more uniform and dispersed, reflecting less attention from the model to specific regions.

Figure 8 shows that the output gradients were calculated with respect to each input pixel, revealing the areas that caused the greatest changes in the classification probability.

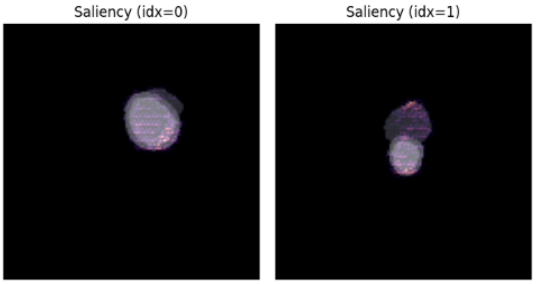

Fig. 8. *Salience Maps*

These visualizations highlighted the heart contours and ventricular borders, indicating that the model based its decisions on the shape and boundaries of the myocardium. The observed pattern confirmed that the CNN learned texture and geometry features, but it could not fully discriminate between healthy and diseased tissues, as many borders appeared prominent even in sections without pathology.

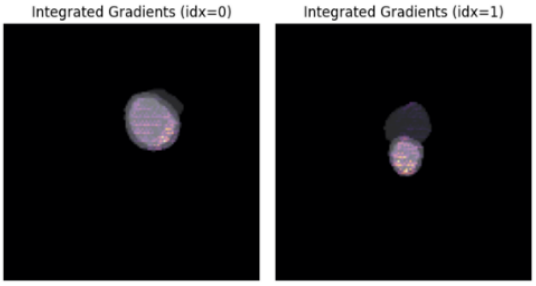

Fig. 9. *Integrated gradients*

The integrated gradient technique, as shown in Fig. 9, offers a more stable perspective, reducing the noise characteristic of pure gradients.

The results showed moderate-intensity concentrations over the mid-septal myocardium and ventricular walls, which coincided with the areas that Grad-CAM identified as most relevant. This agreement between the methods reinforces the confidence that the CNN is indeed learning anatomically plausible patterns, even though its overall accuracy is limited.

4.4.3. YOLOv8

This model, originally designed for object detection, was adapted for the binary classification of heart disease. Owing to its fully convolutional nature and the presence of feature pyramid and self-attention layers, its internal behavior is less intuitive than that of a traditional CNN. Therefore, two spatial explainability methods were applied: Grad-CAM and Occlusion Sensitivity, to analyze the anatomical regions that contribute to the final prediction and validate the consistency of their activations.

The Grad-CAM maps generated from the most representative axial slices of each patient (see Fig. 10) showed intense activation in the ventricular cavities and interventricular septum in the cases classified as diseased cases.

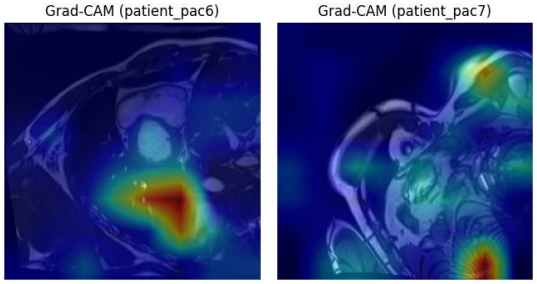

Fig. 10. *Grad-CAM maps (pac6 and pac7)*

In particular, patients pac6 and pac7 showed hot spots (red and yellow) concentrated in the left ventricle and septal wall, consistent with the typical location of abnormalities in dilated or hypertrophic cardiomyopathy.

In the case of patient pac27 (Fig. 11), the Grad-CAM map showed activations slightly shifted toward peripheral areas, reflecting a prediction with moderate confidence (probability = 0.61).

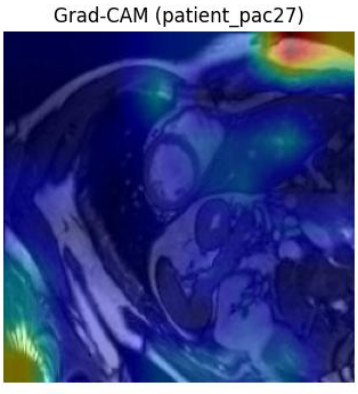

Fig. 11. *Mapas Grad-CAM(pac27)*

This suggests that although the model recognizes pathological signals, its attention may be dispersed in cases with less pronounced cardiac textures or low contrast. In healthy patients, Grad-CAM

activations were diffusely distributed without clear concentrations, coinciding with the absence of anomalous structural patterns.

To complement the analysis, the Occlusion Sensitivity method was applied (see Fig. 12), which measures the decrease in probability when local patches of the image are masked.

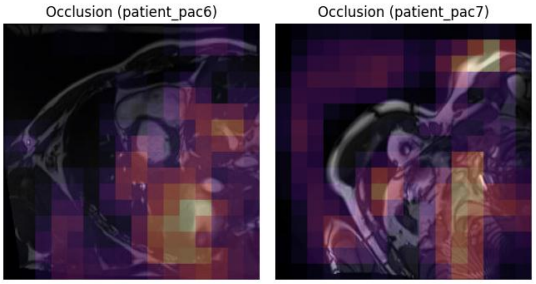

Fig. 12. *Occlusion Sensitivity (pac6, pac7)*

The results showed that the regions whose occlusion caused the greatest decrease in probability (yellow and red areas) coincided with the same areas highlighted by Grad-CAM: the ventricular walls and interventricular septum. In patients with disease, left ventricular occlusion reduced the probability of the "disease" class by more than 40%, confirming that these regions are critically important determinants for the model's decision. In contrast, in healthy patients, occlusion did not produce significant changes in the output, validating the specificity of care.

5. CONCLUSIONS

This study systematically compared three representative artificial intelligence approaches applied to the detection of heart disease using cardiac magnetic resonance imaging: classical machine learning (Random Forest), convolutional neural networks (CNNs), and optimized classification models (YOLOv8).

The methodology encompassed image preprocessing, feature extraction, and performance evaluation using standardized metrics (accuracy, precision, recall, F1-score, and AUC). This was complemented by an explainability analysis (XAI), which allowed for the interpretation of each model's decisions.

The results showed that YOLOv8 achieved the best overall performance, with an accuracy and precision of 80 %, an F1-score of 0.89, and an AUC of 0.83. This reflects a good ability to discriminate between healthy and diseased patients. Its architecture allows the preservation of relevant spatial information and detection of complex structural patterns in the myocardium, demonstrating its potential for critical medical tasks. On the other hand, Random Forest also achieved competitive performance (accuracy = 0.80; recall = 1.00), relying on textural descriptors derived from GLCM matrices, while CNN performed worse (AUC = 0.4167), showing difficulty in correctly generalizing the negative class owing to data imbalance.

From the perspective of explainable artificial intelligence, all three approaches demonstrated physiological coherence in their decisions. Random forest-based classification was based on the homogeneous textural properties of myocardial tissue; CNN partially focused its attention on the left ventricle and the interventricular septum; and YOLOv8 concentrated its activations on anatomically relevant regions, confirming its spatial representation capacity. These findings underscore that explainability not only provides transparency but is also a complementary clinical validation tool for quantitative performance.

In the clinical setting, the high sensitivity observed, especially in YOLOv8 and Random Forest, is encouraging for the development of diagnostic support systems. However, the presence of false positives in deep learning-based models highlights the need to recalibrate decision thresholds, improve class balance, and incorporate cross-validation with specialists.

Finally, this study demonstrates that artificial intelligence is a viable and robust approach for the assisted diagnosis of structural heart disease. Models such as YOLOv8 combine performance and explainability, offering a solid foundation for their implementation in real-world clinical workflows. Future research should focus on increasing data diversity, optimizing interpretability, and evaluating the multicenter reproducibility of these models to consolidate their role as reliable and transparent clinical support tools.

ACKNOWLEDGMENTS

This project was funded by the Universidad de Córdoba, through internal call resources with the project code FI-01-24.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Ministerio de Salud y Protección Social, "Enfermedades cardiovasculares." [Online]. Available: https://www.minsalud.gov.co/salud/Paginas/Enfermedades-cardiovasculares.aspx

[3] Ministerio de Salud y Protección Social, "Minsalud conmemora el Día Mundial del Corazón," Sep. 29, 2022. [Online]. Available: https://www.minsalud.gov.co/Paginas/Minsalud-conmemora-el-dia-mundial-del-Corazon.aspx

[4] Ministerio de Salud y Protección Social, "Plan Nacional de Salud Pública 2022–2031." [Online]. Available: https://www.minsalud.gov.co/plandecenal/Paginas/PDSP-2022-2031.aspx

[5] D. De Santis, A. Trivisonno, and E. G. Caiani, "Accelerated deep learning-based function assessment in cardiovascular magnetic resonance: a clinical comparison with conventional cine sequences," La Radiologia Medica, 2025, doi: https://doi.org/10.1007/s11547-025-02019-6

[6] A. Argentiero, G. Muscogiuri, M. G. Rabbat, C. Martini, N. Soldato, P. Basile, et al., "The applications of artificial intelligence in cardiovascular magnetic resonance—A comprehensive review," Journal of Clinical Medicine, vol. 11, no. 10, p. 2866, 2022, doi: https://doi.org/10.3390/jcm11102866

[7] O. Bernard et al., "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?" IEEE Trans. Med. Imaging, vol. 37, no. 11, pp. 2514–2525, 2018. https://doi.org/10.1109/TMI.2018.2837502.

[8] M. Alsharqi and E. R. Edelman, "Artificial Intelligence in Cardiovascular Imaging and Interventional Cardiology: Emerging Trends and Clinical Implications," J. Soc. Cardiovasc. Angiogr. Interv., vol. 4, no. 3, p. 102558, 2025. https://doi.org/10.1016/j.jscai.2024.102558

[9] Y.-R. Wang, K. Yang, Y. Wen, P. Wang, Y. Hu, Y. Lai, Y. Wang, K. Zhao, S. Tang, A. Zhang, H. Zhan, M. Lu, X. Chen, S. Yang, Z. Dong, Y. Wang, H. Liu, L. Zhao, L. Huang, Y. Li, L. Wu, Z. Chen, Y. Luo, D. Liu, et al., "Screening and diagnosis of cardiovascular disease using artificial intelligence-enabled cardiac magnetic resonance imaging," Nature Medicine, vol. 30, pp. 1471–1480, 2024, doi: https://doi.org/10.1038/s41591-024-02971-2

[10] R. Jafari, R. Azad, and S. Minaee, "Automatic diagnosis of myocarditis disease in cardiac MRI modality using deep transformers and explainable artificial intelligence," arXiv preprint arXiv:2210.14611, 2022. https://doi.org/10.48550/arXiv.2210.14611

[11] D. Li, H. Zhang, and C. Wu, "An algorithm for cardiac disease detection based on magnetic resonance imaging using an improved SA-YOLO model," Scientific Reports, vol. 15, 88567, 2025, doi: https://doi.org/10.1038/s41598-025-88567-3

[12] M. U. Rehman, S. Akbar, and A. Imran, "Predicting coronary heart disease with advanced machine learning classifiers for improved cardiovascular risk assessment," Scientific Reports, vol. 15, 96437, 2025, doi: https://doi.org/10.1038/s41598-025-96437-1

[13] T. Tsampras, T. Karamanidou, G. Papanastasiou, and T. G. Stavropoulos, "Deep learning for cardiac imaging: Focus on myocardial diseases: A narrative review," Hellenic Journal of Cardiology, 2024. https://doi.org/10.1016/j.hjc.2024.12.002

[14] F. Cheng et al., "Learning directional feature maps for cardiac MRI segmentation," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), Cham, Switzerland: Springer, Sep. 2020, pp. 108–117. https://doi.org/10.1007/978-3-030-59719-1_11

[15] D. M. Alsekait et al., "Heart-Net: A multi-modal deep learning approach for diagnosing cardiovascular diseases," Comput. Mater. Contin., vol. 80, no. 3, 2024. https://doi.org/10.32604/cmc.2024.054591

[16] R. K. Sahu, D. R. Sahu, and B. K. Pattanayak, "A hybrid segmentation and classification CAD framework for myocardial infarction detection using cardiac MRI," Scientific Reports, vol. 15, 98893, 2025, doi: https://doi.org/10.1038/s41598-025-98893-1

[17] X. Liu, F. Xing, H. K. Gaggin, C. C. J. Kuo, G. El Fakhri, and J. Woo, "Successive subspace learning for cardiac disease classification with two-phase deformation fields from cine MRI," in Proc. 2023 IEEE 20th Int. Symp. Biomed. Imaging (ISBI), Apr. 2023, pp. 1–5. https://doi.org/10.1109/ISBI53787.2023.10230746

[18] J. Xing, S. Wang, K. C. Bilchick, A. R. Patel, and M. Zhang, "Joint deep learning for improved myocardial scar detection from cardiac MRI," in Proc. 2023 IEEE 20th Int.

Symp. Biomed. Imaging (ISBI), Apr. 2023, pp. 1–5. https://doi.org/10.1109/ISBI53787.2023.10230541

[19] G. P. Veldhuizen, T. Lenz, D. Cifci, M. van Treeck, J. Clusmann, Y. Chen, C. V. Schneider, T. Luedde, P. W. de Leeuw, A. El-Armouche, D. Truhn, and J. N. Kather, "Deep learning can predict cardiovascular events from liver imaging," JHEP Reports, 2025, Art. no. 101427. https://doi.org/10.1016/j.jhepr.2025.101427

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," International Journal of Computer Vision, vol. 128, pp. 336–359, 2019, https://doi.org/10.1007/s11263-019-01228-7

[21] J. Liu, A. Wei, L. Cao, X. He, and C. Tang, "Contrastive trustworthy prototype learning for multi-modality myocardial pathology segmentation," Applied Soft Computing, vol. 173, Art. no. 112909, 2025, doi: https://doi.org/10.1016/j.asoc.2025.112909

[22] Z. Dong, Y. Tang, P. Sun, G. Yin, K. Zhao, X. Ma, and S. Zhao, "Early identification of myocardial microstructural alterations in hypertrophic cardiomyopathy with in vivo cardiac diffusion-tensor imaging," Radiology: Cardiothoracic Imaging, vol. 7, no. 1, Art. no. e240009, 2025, doi: https://doi.org/10.1148/ryct.240009