Magnetic Resonance Imaging Analysis for Cardiomyopathy Classification with Attention Mechanisms Support

Diogo Santiago, Leila Cristina C. Bergamasco

Abstract— **The increasing availability of medical imaging exams, such as magnetic resonance imaging, generates a large volume of data, making its analysis complex and challenging. In this scenario, advanced computational approaches can optimize the interpretation of these images and assist in the early diagnosis of cardiovascular diseases. This work aims to unify contemporary approaches in the evaluation of cardiomyopathy. With the support of radiomic analysis, which extracts information from statistical and texture characteristics of a medical image, and features derived from a classical neural network for computer vision, such as ResNet50, promising results can combined and obtained [3]. Some forms of attention mechanisms are applied, including self-attention and selective attention using Squeeze and Excitation Nets. The results confirm that the combination of information from various domains regarding a given patient, when integrated, can lead to more interesting outcomes compared to analyzing data in isolation. This study aims to apply the approaches mentioned above, based on previous literature, in an innovative application for cardiomyopathy testing.**

Index Terms— **Radiomics, Attention Mechanism, Transformers, Cardiomyopathy, Medical Imaging.**

I. INTRODUCTION

SINCE the early 2000s, the volume of data generated in medicine has grown exponentially,this rapid growth highlights the need for tools capable of efficiently processing and analyzing vast amounts of information.

Medical imaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), have become essential in modern medicine, offering detailed three-dimensional representations of human anatomy. These imaging techniques not only improve diagnostic accuracy but also generate large datasets that can be quantitatively analyzed. Concurrently, artificial intelligence (AI) has driven significant advancements in diagnostic imaging, enhancing efficiency and precision in medical decision-making.

Deep neural networks have demonstrated high performance in computer vision tasks, including image classification, object detection, and segmentation. Once optimized on a given dataset, these models can extract discriminative features that enhance medical image analysis. Furthermore, transformer architectures, commonly used in autoregressive generative models such as large language models (LLMs), have gained prominence due to their parallelization capabilities and self-attention mechanisms, which allow models to focus on the most relevant parts of the input data.

Additionally, image processing techniques such as texture analysis have been employed for decades in various medical domains. Radiomics, a powerful approach for extracting quantitative features from medical images, captures patterns that may not be perceptible to the human eye. This methodology has shown potential applications in fields such as oncology and cardiology [8].

In cardiac imaging analysis, texture-based MRI assessments have been used to evaluate the risk of post-myocardial infarction arrhythmia. Specifically, texture analysis of contrast-enhanced and non-contrast MRI scans in cardiomyopathy patients has been explored for predicting clinical outcomes [10]. Among cardiomyopathies, *hypertrophic cardiomyopathy* (HCM) is one of the most prevalent frequently diagnosed in young and middle-aged individuals. While often asymptomatic, HCM can lead to severe conditions such as heart failure and stroke, making early diagnosis crucial in preventing adverse outcomes. In this context, radiomics—which extracts high-dimensional data from medical images, often including texture analysis—holds promise for early diagnosis and risk assessment in cardiomyopathy [5].

Integrating AI and radiomics represents a promising strategy for detecting cardiomyopathies and other cardiac conditions. Recent studies have demonstrated that combining deep features with radiomic characteristics can significantly improve the predictive performance of diagnostic models for lung cancer using CT imaging. Models leveraging self-attention mechanisms to analyze concatenated data have achieved up to 82.35% accuracy, and an AUC of 0.74 [6].

Thus, this study proposes the implementation and validation of a fusion strategy that combines radiomic and deep learning features with self-attention mechanisms to improve cardiomyopathy classification. Beyond advancing the state-of-the-art in medical diagnosis, this research aims to contribute to the development of more effective and accessible solutions for clinical decision support [2].

II. DATASET

Two datasets were used in this study: ACDC and Sunny-Brook. Both are publicly available and intended for research

purposes. The ACDC dataset consists of 150 images, with 100 for training and 50 for testing, evenly distributed across five classes: DCM, HCM, NOR, MINF, and RV. The SunnyBrook dataset contains 45 images, and to maintain proportional compatibility with ACDC, it was divided into 30 images for training and 15 for testing. The dataset includes the following classes: NOR, IC, IC-I, and HIP. Details about each class are provided in the following sections dedicated to each dataset.

A. ACDC Dataset

The ACDC dataset was created using real clinical exams obtained from the University Hospital of Dijon (France). The acquired data was fully anonymized and processed in compliance with the regulations established by the local ethics committee of the Dijon hospital.

This dataset covers several well-defined pathologies, with a sufficient number of cases to: (1) properly train machine learning methods and (2) clearly evaluate variations in key physiological parameters derived from cine-MRI, particularly diastolic volume and ejection fraction.

The dataset consists of 150 for exams, 100 for training, and 50 for tests. Each exam is from a different patient , and the dataset is divided into five equally distributed subgroups. The five distinct classes are: *dilated cardiomyopathy* (DCM), HCM, *normal* (NOR), *myocardial infarction* (MINF), and *right Ventricular abnormality* (RV). The DCM and HCM classes are interpreted as indicative of cardiomyopathy, while NOR, MINF, and RV represent normal heart conditions. Details on these classes can be found in Table I.

Additionally, the dataset includes segmentation masks, allowing for potential segmentation applications. The label values range from 0 to 3, representing voxels associated with the background (0), right ventricle cavity (1), myocardium (2), and left ventricle cavity (3). Figures 17, 18, and 19 illustrate the images and their respective segmentation masks for DCM, HCM, and NOR. These masks display three distinct shades of gray corresponding to the described segmentation.

TABLE I
ACDC LABELS

Group	Quantity	W/ Cardiomiopaty	W/O Cardiomiopaty
NOR	30	0	30
DCM	30	30	0
HCM	30	30	0
MINF	30	0	30
RV	30	0	30
Total:	150	60	90

ACDC - Redistribution for normal $\times$ cardiomyopathy classification.

B. SunnyBrook Dataset

The Sunny dataset, also known as the 2009 Cardiac MR Left Ventricle Segmentation Challenge data, consists of 45 MRI images from a mixed group of patients and pathologies, including healthy individuals, hypertrophy, heart failure with infarction, and heart failure without infarction. There are four pathological groups in this dataset, classified as follows:

a) Heart failure with infarction (IC-I): Group with an ejection fraction (EF) ¡ 40% and evidence of late gadolinium enhancement (Gd). b) Heart failure without infarction (IC): Group with EF ¡ 40% and no late gadolinium enhancement. c) Left ventricular hypertrophy (HIP): Group with normal EF ($> 55\%$) and a left ventricular mass-to-body surface area ratio > 83 g/m². d) Normal (NOR): Group with EF $> 55\%$ and no hypertrophy.

For this study, the HIP and NOR classes were used to classify cases of cardiomyopathy and normal conditions, respectively. Table 6 shows the distribution of the 45 patients and the cardiovascular activity values used for patient classification.

III. METHODOLOGY

In this section, first introduce the image preprocessing and feature extraction process, which involves selecting and extracting features from MRI images. Then, we introduce the base model, originally applied to non-small cell lung cancer (NSCLC) early recurrence prediction. And last, we describe the design of our proposed architecture with a new concatenation, using segmented masks, and in addition to using the self-attention mechanism, it also used selective attention in the form of *squeeze and excitation* (SE) Net modules.

Studies by [4] demonstrated that network performance could be enhanced by explicitly incorporating learning mechanisms that help capture spatial correlations without requiring additional supervision. In this context, the SE mechanism was proposed to allow the network to recalibrate features, enabling it to learn how to use global information to selectively emphasize informative features while suppressing less useful ones [7].

The change in the concatenation method was intentional, as the adapted version includes a new convolutional layer that is not present in the original version, aiming to extract relevant information from the fusion of embeddings. The adapted version also adds a selective attention layer as an SE block, which precedes the self-attention block [9].

Another modification in the adapted version is that the self-attention block is applied N times instead of just once, as in the base work, leading to a noticeable improvement in results. It is important to note that both attention mechanisms serve different purposes: while the selective attention block focuses on the relationships between the embedding channels, the self-attention block efficiently optimizes and weights the fused feature vector.

The intuition behind applying attention mechanisms with different purposes is inspired by the work of [12].

A. Data Preprocessing

For the preprocessing step, the ACDC dataset was used. The ACDC dataset originally contains five distinct classes and we organized it for a binary cadyomiopathy classification. The set of frames used are defined in the diastolic phase, and the number of frames varies per patient. The radiomic features are extracted using the PyRadiomics library. This tool allows 3D cardiac image volumes and their respective masks to be used as input, extracting manual features related to texture, shape, grayscale intensity, etc.. The result is a vector of 78 radiomic features.

Deep features are extracted using a ResNet50 network, excluding the final classification layer. This process is applied to each slice of the 3D volume for both the cardiac image and its corresponding mask, resulting in feature vectors of size $100,352$. Following the base model's methodology, an F-Test is applied to reduce the dimensionality from deep features from 100,352 to 27,372 and radiomic features from 78 to $EMBED_{size} \in \{24, 48, 64\}$.

The $EMBED_{size}$ value varies in the adapted versions, serving as a hyperparameter, ensuring that all embeddings have the same size. Feature concatenation is performed In the last dimension for the original model and its variations and in the first dimension for adapted versions. It is important to note that, unlike the base model version, which contains only one deep feature vector, the adapted versions may include an additional vector corresponding to the mask with the region of interest.

B. Base Model

The work from Ai et al. (2023) proposed deep learning based on a self-attention mechanism for NSCLC early recurrence prediction. Firstly, radiomics was applied using diverse machine learning techniques to extract handcrafted features from CT images, encompassing texture, shape, grayscale, etc. Radiomics-based methods mainly rely on handcrafted design to extract numerous quantitative features from CT images, including tumor shape, size, density, texture, edge, etc [1]

Subsequently, a pre-trained ResNet50 network was utilized to extract deep features that encapsulate high-level semantic and representation information from the CT images. These features were then fused into a unified feature vector, and a self-attention module was applied, ending with a classification layer for the NSCLC early recurrence prediction [1]. The architecture can be seen in Figure 1.
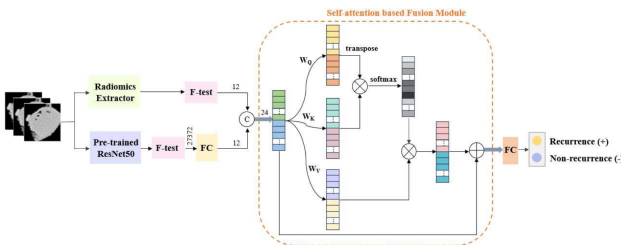

Fig. 1.   Proposed Architecture for NSCLC early recurrence prediction. First, radiomic and deep features are extracted, an F1-Test reduces their dimensions are concatenated and sent to a self-attention module.

C. Proposed Model

The schematic of the proposed architecture is shown in Figure 2. Deep features are extracted from the images and masks using ResNet50 and undergo a linear transformation to a size of $27,372$, as per the base model. Radiomic features are extracted using various statistical models, resulting in 78 features.

An F-Test set is applied to the feature vectors to ensure they have the same size, which is determined by the $EMBED_{size}$

parameter. The selected features are then concatenated and passed through the convolution module, which consists of a 1D convolution, an SE block, and another 1D convolution. The first convolution converts the number of channels to 16, the SE block maintains the dimensionality, and the second convolution reduces the channels from 16 to 1. The process continues through the self-attention module, and finally, a linear layer with a single neuron performs binary classification.

D. Selective Attention Module

Selective attention plays a crucial role in improving the performance of deep learning models by allowing them to focus on the most relevant features in their internal representations. Specifically, the SE block is designed to adaptively recalibrate feature channels by assigning higher weights to the most informative characteristics. This mechanism operates in two main steps: first, it reduces the dimensionality of global features (squeeze) to capture overall context, and then it applies an excitation mechanism to adjust the weights of each channel based on their relative importance. By integrating SE blocks into an architecture, the model enhances its ability to discriminate relevant patterns while suppressing noise or irrelevant information, leading to greater accuracy and efficiency in data processing [12].

Since the SE block is added only in the adapted version to improve results, the resulting concatenated vector is 3×64, assuming $EMBED_{size} = 64$ and the inclusion of the mask. A 1D convolutional layer is applied with 16 channels, a kernel size of 3, a stride of 1, and padding of 1 to maintain the original dimensions after processing [4].

With 16 channels as the output of the first convolution, the SE block is applied to identify the most relevant channels and rescale the original input. The SE block has a reduction parameter r, which was set to 16, as suggested by the authors of the original study. Finally, another 1D convolution is applied with a single output channel, resulting in a one-dimensional vector, which serves as input for the next block, the self-attention module.

E. Self-Attention Module

The self-attention mechanism was employed to learn the importance of each feature and capture long-range dependencies. As illustrated in Figure 2, the self-attention module is used to map a query (Q), key (K), and value (V) into an attention-weighted value.

The model uses 24 concatenated features F_{hd} in the original version and $EMBED_{size}$ features in the adapted versions. From the resulting vector, each feature is projected onto three learnable matrices: key matrix (K), query matrix (Q), and value matrix (V) through dot-product operations with their corresponding weight matrices W_Q, W_K, and W_V. Consequently, the values Q, K, and V are defined as $W_Q F_{gd}$, $W_K F_{gd}$, and $W_V F_{gd}$, respectively, where W_Q, W_K, and W_V represent the linear transformations applied to generate the matrices Q, K, and V [11].

The self-attention module is defined in Equation 1, where d_k represents the dimension of K. Without requiring recurrent or
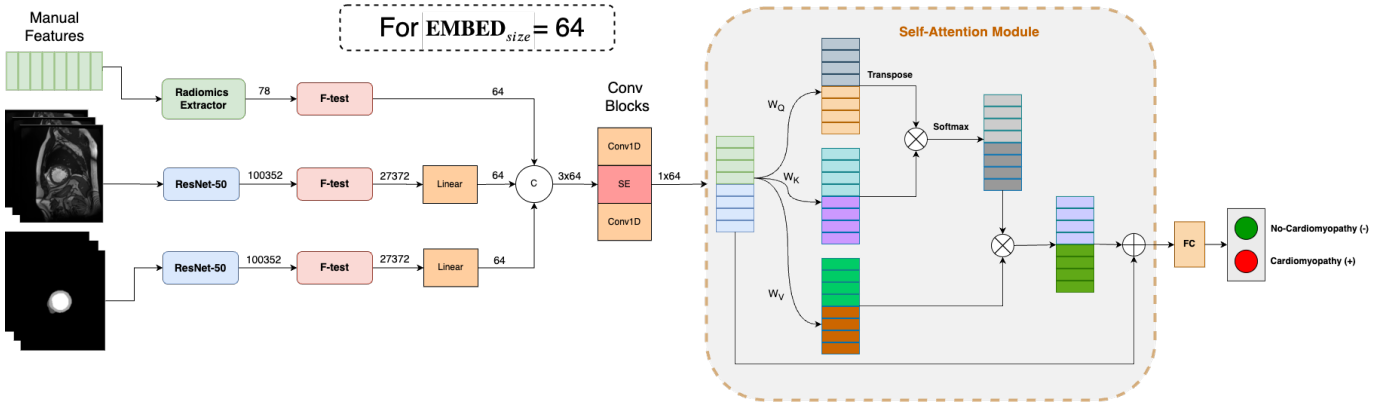
Fig. 2.   Proposed Architecture for cardiomyopathy classification. The concatenation is made in the channel dimension. The major difference is the addition of the heart mask and the SE Module.

convolutional operations, the self-attention module can model long-term dependencies between input features. This module adaptively computes feature weights based on their importance and relevance, enhancing the associations between radiomic and deep features [11].

This process strengthens the expressive power of the fused features and enables the model to focus more precisely on the most informative features, assisting in medical image diagnosis while reducing the influence of irrelevant features in prediction. Additionally, the model can dynamically allocate attention to different cardiac MRI (MRI) image samples, allowing for better adaptation to the feature representations of different cases, improving both accuracy and generalization.

The loss function used in the model is the binary cross-entropy denoted by the Equation 2. Here, $\mathcal{L}_{bce}$ denotes *binary cross entropy* (BCE), N represents the number of cardiac MRI, r denotes the target class of cardiomyopathy, and $\hat{r}$ is the predicted value by the model. A value of 1 indicates the presence of cardiomyopathy, while 0 indicates its absence.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \qquad (1)$$

$$\mathcal{L}_{bce} = -\frac{1}{N}\sum_{i=1}^{N}(r_i \ln \hat{r}_i + (1 - r_i)\ln(1 - \hat{r}_i)) \qquad (2)$$

IV. EXPERIMENTS AND RESULTS

This chapter presents the results of the proof of concept conducted with the developed algorithms, including both the baseline model and the adapted models, with the goal of comparing them to the objectives outlined in this dissertation. The tests performed aim to evaluate the algorithm's efficiency under the various adaptations discussed.

A. Experiments Base Model

A proof of concept was conducted on the ACDC dataset using the baseline model for an initial evaluation, which was also used as the baseline reference. The baseline model was implemented according to the original paper, and its implementation is shown in Figure 1.

The training dataset consists of 100 patient exams, considering only image slices from the diastolic phase. First-order features were extracted using the *PyRadiomics* library, resulting in 78 values that compose the radiomic features.

For deep feature extraction, a frozen ResNet50 network was used, excluding its final linear layer, which was originally responsible for classifying 1,000 classes from the ImageNet dataset. The final output was a feature vector containing 100,352 deep features.

In this experiment, the model was trained using binary cross-entropy as the objective function, a learning rate of 0.0001, the Adam optimizer, a batch size of 1, and approximately 60 epochs. During training, it was observed that the error stabilized over time. Additionally, a data randomization strategy was employed during training. Finally, the sigmoid function was applied to the output vector of the model so we a probability result between 0 and 1. For the classification, it was considered that values above 0,5 are considered cardiomyopathy and value below or equal 0,5 are considered normal cases.

B. Experiments Adapted Model

The adapted models represent modifications derived from the baseline model, aiming to evaluate whether changes either in hyperparameters or in architectural components can yield promising improvements over the baseline results. These modifications can be categorized as follows:

1) Changes to EMBED_{size}, where the values $[24, 48, 64]$ were tested in an attempt to preserve more of the original information.

2) In the adapted versions, the respective masks are also incorporated and later concatenated with the radiomic and deep features.

3) The concatenation process differs from the original model. For instance, assuming $\text{EMBED}_{size} = 12$, the original version produces a 1×24 output, whereas in the adapted version, concatenation is performed along the first dimension, resulting in a 2×12 output. This adjustment is made because the adapted model introduces a new convolutional block, consisting of a convolution operation and SE blocks.

The remaining architecture follows the original design, except that the self-attention block can be applied N times. This means its output is recursively fed back as input N times, making N a tunable hyperparameter in the experiment. The tested values for N were $[1, 2, 4, 6]$. These changes do not alter the overall structure of the initial architecture but allow the self-attention module to be executed N times, as proposed in [11] and seen in Figure 3.
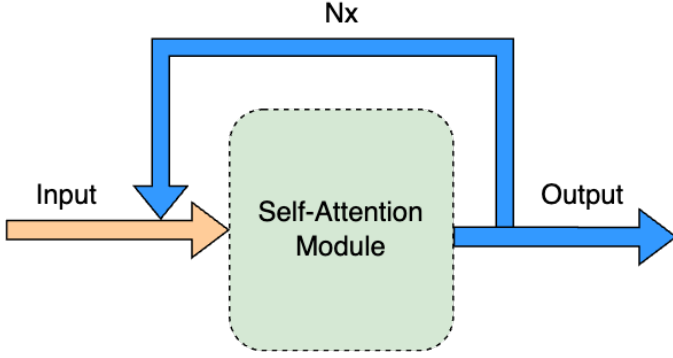

Fig. 3. Self-attention module recurrence.

C. Results Base Model

With the preprocessed and previously stored data, training is performed using the radiomic and deep feature vectors. Partial metrics were saved in real-time during training. For the baseline model, the hyperparameters are: $EMBED_{size} = 12$, $N = 1$ (i.e., a single self-attention block), concatenation dimension = 1, learning rate = $0,0001$, optimizer Adam, batch size = 1 and approximately 60 epochs.

Once trained, inference was performed on the test dataset, applying the sigmoid function and setting 0.5 as the decision threshold, where values above this threshold are classified as cardiomyopathy. In contrast, the others are classified as normal.

The resulting metrics can be found in Table II. The confusion matrix is shown in Figure 4.

TABLE II
EXPERIMENT METRICA - BASE MODEL

Metric	Value
Accuracy	0.58
Precision	0.47
Recall	0.45
AUC	0.55

D. Results Adapted Model

For the adapted versions, the objective was to modify both the hyperparameters and the model architecture to compare the results with the baseline version. The first experiments were conducted by varying the $EMBED_{size}$, and the self-attention block was applied N times instead of just once. These experiments were executed by generating combinations of these hyperparameters to cover as many scenarios as possible within the proposed methodology.
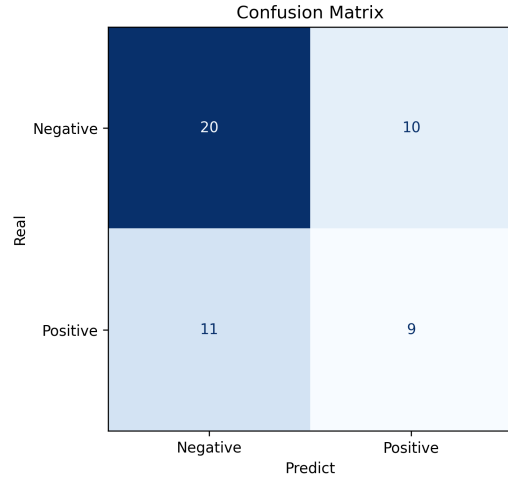

Fig. 4. Confusion Matrix - Base Model.

Additionally, a convolutional module was introduced in the adapted versions, preceding the self-attention module. This convolutional module consists of three blocks: a 1D convolutional layer, an SE block, and another 1D convolutional block. With the option to add a third feature vector, represented by the mask, it became possible to identify intrinsic relationships involving this additional vector. The concatenation is now performed along the first dimension, making each feature vector a separate channel.

In the convolutional module, the first 1D convolutional layer increases the number of channels to 16, meaning that 16 filters are applied for spatial information extraction. The SE block represents the selective attention layer, which assigns importance to the most relevant channels by identifying their significance and scaling them in the original input using an optimized scalar value learned during training. Further details can be seen in Figure 5. In this project, the SE block is adapted for 1D applications, unlike the original study, which focuses on its use in well-established computer vision models. The reduction factor r in the SE block was fixed at 16, as empirical tests conducted by the original study's authors confirmed that this value yields the best results.

As we already have the base model results, Table III shows the experiments that include the mask as a third feature vector, and Table IV presents the addition of convolutional and SE blocks.

Several observations can be made from the experiment. The training process almost always exhibits overfitting, reaching nearly 100% accuracy. The limited number of training samples available can justify this. Another notable point is that larger feature vector sizes (Emb) tend to yield better results. Additionally, increasing the number of self-attention blocks (N_Attn) also improves performance, except when set to 6. This is likely due to the model becoming overly complex compared to the available training data, negatively impacting its generalization ability.

The best results were obtained in versions that incorporated the SE block and used four self-attention blocks, with the best experiment achieving 78% accuracy.
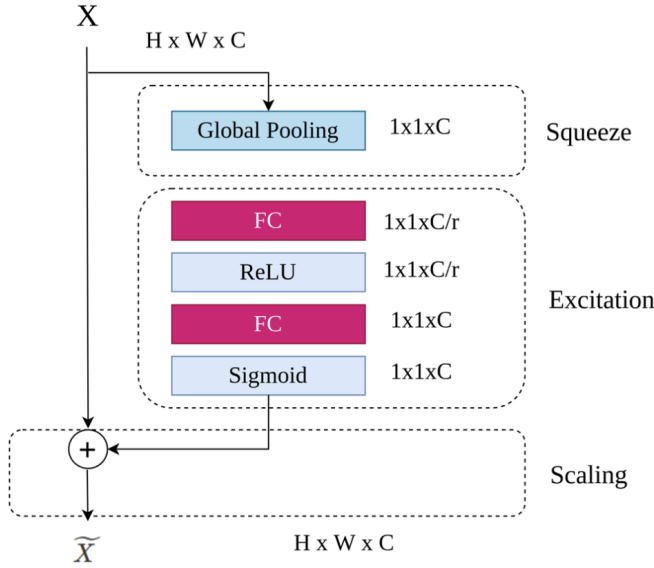
X

H x W x C

Global Pooling 1x1xC   Squeeze

FC 1x1xC/r
ReLU 1x1xC/r   Excitation
FC 1x1xC
Sigmoid 1x1xC

+   Scaling

$\widetilde{X}$ H x W x C

Fig. 5. SE Block Composition

TABLE III
ACDC METRICS - ADAPTED MODELS + MASKS
BOLD MEANS BETTER ACC.

Emb	N_Attn	Acc	Precision	Recall	F1	AUC
24	1	0.44	0.35	0.45	0.39	0.44
24	2	0.46	0.38	0.55	0.45	0.48
24	4	0.48	0.36	0.40	0.38	0.47
24	6	0.48	0.38	0.45	0.41	0.47
48	1	0.52	0.44	0.70	0.54	0.55
48	2	0.58	0.48	0.55	0.51	0.57
48	4	0.62	0.52	0.60	0.56	0.62
48	6	0.60	0.50	0.50	0.50	0.58
64	1	0.58	0.48	0.60	0.53	0.58
64	2	0.56	0.45	0.50	0.48	0.55
64	**4**	**0.64**	**0.54**	**0.65**	**0.59**	**0.64**
64	6	0.62	0.53	0.45	0.49	0.59

TABLE IV
ACDC METRICS - ADAPTED MODELS + MASKS + SE BLOCKS
BOLD MEANS BETTER ACC.

Emb	N_Attn	Acc	Precision	Recall	F1	AUC
24	1	0.66	0.56	0.70	0.62	0.67
24	2	0.68	0.58	0.75	0.65	0.69
24	4	0.70	0.59	0.80	0.68	0.72
24	6	0.68	0.59	0.65	0.62	0.67
48	1	0.70	0.58	0.90	0.71	0.73
48	2	0.70	0.62	0.65	0.63	0.69
48	4	0.74	0.65	0.75	0.70	0.74
48	6	0.70	0.60	0.75	0.67	0.71
64	1	0.64	0.54	0.65	0.59	0.64
64	2	0.72	0.64	0.70	0.67	0.72
64	**4**	**0.78**	**0.76**	**0.65**	**0.70**	**0.76**
64	6	0.76	0.65	0.85	0.74	0.77

evant information in the feature map. This technique not only improved the model's performance but also provided a more intuitive way to explain AI decisions, a crucial aspect for its acceptance in clinical practice.

Although the results are promising, this study also revealed limitations, such as the need for more robust and diverse datasets and the challenge of interpreting deep learning models, a recurring issue in medical AI applications.

For future work, it is suggested to explore the combination of medical images with clinical and genetic data to create multimodal models capable of providing more holistic and personalized diagnoses.

ACKNOWLEDGMENT

REFERENCES AND FOOTNOTES

A. References

REFERENCES

[1] Yang Ai, Yinhao Li, Rahul Kumar Jain, and Yen-Wei Chen. A self-attention based fusion model of radiomics and deep features for early recurrence prediction in nsclc. In *2023 IEEE 12th Global Conference on Consumer Electronics (GCCE)*, pages 833–837. IEEE, 2023.

[2] Jay Brieler, Matthew A Breeden, Jane Tucker, Saint Louis University, and St Louis. Cardiomyopathy: An Overview. 96(10).

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition.

[4] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-Excitation Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141.

[5] Soonil Kwon, Hyung-Kwan Kim, Bongseong Kim, Hyun-Jung Lee, Kyung-Do Han, In-Chang Hwang, Yeonyee E. Yoon, Jun-Bean Park, Heesun Lee, Seung-Pyo Lee, Goo-Yeong Cho, and Yong-Jin Kim. Comparison of mortality and cause of death between adults with and without hypertrophic cardiomyopathy. 12(1):6386.

[6] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G.P.M. Van Stiphout, Patrick Granton, Catharina M.L. Zegers, Robert Gillies, Ronald Boellard, André Dekker, and Hugo J.W.L. Aerts. Radiomics: Extracting more information from medical images using advanced feature analysis. 48(4):441–446.

[7] Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, R. Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten Digit Recognition with a Back-Propagation Network. In *Advances in Neural Information Processing Systems*, volume 2. Morgan-Kaufmann.

V. CONCLUSION

This study presented an innovative approach for the classification of cardiomyopathies using cardiac MRI images combined with attention mechanisms, promoting the integration of radiomic and deep features. The proposed method was built on strong theoretical foundations, employing modern deep learning techniques, such as the SE Net architecture and self-attention mechanisms, to explore discriminative features in medical images.

The results highlighted the efficiency of the proposed model in identifying patterns associated with cardiomyopathies, surpassing traditional approaches in terms of accuracy and generalization. The fusion of radiomic information, which captures textural and statistical details, with deep features extracted from convolutional networks, proved to be effective in handling the intrinsic complexity of cardiomyopathies. Additionally, the use of attention mechanisms allowed the model to prioritize the most relevant regions of the images, enhancing its expressiveness.

The use of SE and spatial attention in the medical imaging domain represented an innovative strategy by prioritizing rel-

[8] Marius E. Mayerhoefer, Andrzej Materka, Georg Langs, Ida Häggström, Piotr Szczypiński, Peter Gibbs, and Gary Cook. Introduction to Radiomics. 61(4):488–495.

[9] Keiron O'Shea and Ryan Nash. An Introduction to Convolutional Neural Networks.

[10] R. Schofield, B. Ganeshan, M. Fontana, A. Nasis, S. Castelletti, S. Rosmini, T.A. Treibel, C. Manisty, R. Endozo, A. Groves, and J.C. Moon. Texture analysis of cardiovascular magnetic resonance cine images differentiates aetiologies of left ventricular hypertrophy. 74(2):140–149.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need.

[12] Mingzhe Yang, Yuan Zhou, Haiyang Wang, and Shuoshi Li. Neural Network Design Based on Differentiable Attention Module Search. In *2024 43rd Chinese Control Conference (CCC)*, pages 8643–8648.