An XAI-Based Deep Learning Framework for Coronary Artery Disease Diagnosis using SPECT MPI polar map images

1st Do Thanh Ton
Hung Yen University of Technology and Education
Hungyen, Vietnam
thanhtonvk@gmail.com

2nd Nguyen Chi Thanh
Institute of Information Technology, AMST
Hanoi, Vietnam
thanhnc80@gmai.com
Corresponding author

3rd Phung Nhu Hai
Institute of Information Technology, AMST
Hanoi, Vietnam
hainda59@gmail.com

4th Nguyen Van Hau
Hung Yen University of Technology and Education
Hungyen, Vietnam
nvhau66@gmail.com

5th Tran Trung Kien
Institute of Information Technology, AMST
Hanoi, Vietnam
t2kien@gmail.com

6th Trung Nguyen Thanh
Institute of Information Technology, AMST
Hanoi, Vietnam
thanhtrungys@yahoo.com

Abstract*-Our study aimed to develop an explanatory method for predicting Coronary Artery Disease (CAD) classification using spect images. As we all know, deep neural networks usually consist of many layers connected to each other through interlocking network nodes. Even if we check the classes and describe their relationships, it is difficult to understand entirely how active neural networks make predictions. Therefore, deep learning is still considered a "Black box". Existing XAI (eXplainable Artificial Intelligence) approach can provide insights into the inside of a Deep Learning model allowing for transparency and interpretation. Our previous research helps doctors diagnose the CAD of patients by developing deep learning models using a multi-stage transfer learning framework. The model achieved 0.955 accuracy, 0.932 AUC, 0.944 sensitivity, and 0.889 specificity, showing effective performance. Our dataset includes 218 SPECT images from 218 imported patients collected at 108 Hospital in Hanoi, Vietnam. In this paper, We propose an explainable Deep Learning framework using three popular XAI approaches: LIME, GradCam, and RISE. These XAI approaches are effective tools for interpreting the prediction of deep learning models. We evaluate the effectiveness of the interpretation by visualizing the explained regions and using improved deletion and insertion with a threshold limit suitable for Binary Classification. Allows the doctors to make a quick and confident diagnosis using the explanations provided. Predictions present that we have built a good and effective TransferLearning model and explanatory methods to diagnose CAD and provide medical interpretation.

***Index Terms* – XAI, Transfer Learning, Deep Learning, LIME, GradCAM, RISE**

I. INTRODUCTION

Coronary artery disease(CAD) is the most common form of cardiovascular disease in the elderly. It is one of the leading causes of death in the world. In Vietnam, the rate of patients dying from coronary heart failure is relatively high, accounting for 11% - 36%, and is showing signs of increasing yearly.

Nowadays, SPECT-MPI is a remarkably effective method concerning the diagnosis of CAD [9]. This approach provides 3D information about the distribution of radioactive compounds in the heart, reducing the number of unnecessary angiograms and allowing for appropriate treatment planning. MPI is a noninvasive imaging modality where injected radiopharmaceutical uptake is measured using SPECT to diagnose CAD [12]. Research in CAD diagnostic support using SPECT MPI has been carried out a lot. Some studies such as [1] and [2] use Machine Learning (ML) algorithms such as Adaptive Boosting, Gradient Boosting, Random Forests, and Xgboost. Although these methods achieve >90% accuracy, it requires a large amount of data and has to go through many processes before the doctor can conclude whether he has CAD. The emergence of Deep Learning (DL) has recently become more widely used in CAD diagnosis. This trend gradually improves the training time and cost significantly compared to ML. Therefore, several DL-based CAD diagnostic models have been developed and achieved promising results. Papandrianos et al. [2] improved the model using DL. Specifically, the accuracy reaches 94.58%.

Deep learning has made significant progress in image analysis and prediction. Image Analysis was often introduced using a human design system in the past. In CAD diagnosis using SPECT MPI image, the doctor will have to analyze the image and predict whether the patient has the disease or not based on the area of the lesion. In deep learning, this is learned by neural networks to optimize the input. However, neural networks usually consist of many layers that are connected through interlocking network nodes. Even if we examine the classes and describe their relationships, it is difficult to understand how all active neural networks make predictions. Before that, deep learning was still considered a "Black box". It is this that makes it impossible for users to trust the model without knowing how its predictive logic works, whether it is similar to expert diagnostic logic or not. It also cannot explain to the patient why he has CAD. It is for these reasons that eXplainable Artificial Intelligent (XAI) was born to solve those problems.

During the 1970s to 1990s, symbolic reasoning systems, such as MYCIN, [4] GUIDON, [5] SOPHIE, [6] and PROTOS[7][8] were explored that could represent, reason about, and explain their reasoning for diagnostic, instructional, or machine-learning (explanation-based learning) purposes In the 1980s through the early 1990s, truth maintenance systems (TMS) were developed to extend the capabilities of causal-reasoning, rule-based, and logic-based inference systems.

By the 1990s, researchers also began studying whether it is possible to meaningfully extract the non-hand-coded rules being generated by opaque trained neural networks. [9] Researchers in clinical expert systems creating neural network-powered decision support for clinicians have sought to develop dynamic explanations that allow these technologies to be more trusted and trustworthy in practice.

In addition, there has been research to make glass-box models more transparent to inspection. These include decision trees, Bayesian networks, sparse linear models, etc. The Association of Computing Machinery Conference on Fairness, Accountability, and Transparency (ACM FAccT) was established in 2018 to study transparency and explainability in socio-technical systems, many of which include AI.

It can be said that the concept of XAI has been around for a long time, but due to technical limitations, it has not been developed. Recently, when advanced CNN network models and technologies appeared, XAI (eXplainable Artificial Intelligence) was developed and promoted its effectiveness because the more accurate the DL model, the more complex it is and the easier it is. There are many popular XAI methods such as LIME (Local Interpretable Model-agnostic Explanations) [10], SHAP (Shapley Additive exPlanations) [11], Influence Functions [12], Integrated Gradient [13], Grad-Cam [15], RISE (Randomized Input Sampling for Explanation) [14], etc. In this study, we propose an explainable DL Framework using three methods: Grad-Cam, LIME, and RISE. The main contributions of this paper are the following:

- Multiple interpretation methods are used to interpret the model in many respects.
- We evaluate the effectiveness of XAI methods in detail for each label (CAD and NonCAD) on aspects: visualization, an improved deletion algorithm, and an improved insertion algorithm.

The organization of this paper is structured as follows: Section 2 reviews the work related to CAD interpretations: GradCam, RISE, and LIME. Section 3 describes the methodology of the usage model and the XAI methods. Section 4 presents the results of the interpretation. Section 5 compares the effectiveness of the XAI methods, the work involved, and the paper conclusions of the paper.

II. RELATED WORK

Nikolaos I.Papandrianos et al., in [16], used transfer learning with the VGG16 network for CAD classification problems. The study used a dataset that included cases from 625 patients as representative of stress and rest, including 127 infarct, 241 ischemic, and 257 normal cases that were disproved. Result: The model achieves 93.3% accuracy and 94.58% AUC. The explanation for the model using GradCAM-based color visualization.

Liu et al., in [17], aimed to investigate a DL approach to improve the diagnostic accuracy of CAD. A total of 37,243 patients in a stress mode were selected in this study, wherein count profile maps were extracted from SPECT MPI images. In addition, clinical data were included, such as gender, BMI, length, stress type, and radiotracer, along with the options of adding or not adding attenuation correction. A DL methodology with transfer learning was developed. More specifically, ResNet-34 was utilized and compared against a conventional quantitative perfusion defect size (DS). The DL prediction was evaluated utilizing five-fold cross-validation. The AUC results for the DL and DS methods were 0.872 ± 0.002 and 0.838 ± 0.003, respectively, the DL method accordingly showing better performance.

Apostolopoulos et al. in [18] developed a hybrid CNN-random forest model and utilized polar map images and clinical data to classify patients' status into normal and abnormal and further compared the results against nuclear experts' diagnoses. The proposed model was evaluated with the 10-fold cross-validation technique and achieved 79.15% accuracy, 77.36% sensitivity, and 79.25% specificity while exhibiting similar overall results to those provided by experts.

In our previous research [3] – we continue to use the pre-proposed model for this study. We used 218 SPECT images from 218 patients in the Department of Nuclear Medicine of 108 Hospital, Hanoi, Vietnam. With a number of CAD: 112 (51.37%) and a number of non CAD: 106(48.63%). The dataset has been processed and consulted by experts and doctors. We proposed a multi-stage transfer learning framework and evaluated the proposed framework by 15 pre-trained deep CNNs, which are well-trained on ImageNet (more than 1 million images for 1000 classes). Since the dataset is small, they used Global Average Pooling (GAP) instead of Flatten function at the end of the feature extraction to prevent overfitting. Furthermore, the GAP is more native to the convolution structure by enforcing correspondences between feature maps and categories so the feature maps can be easily interpreted as category confidence maps. We used all the most popular pre-trained CNNs such as VGG, Xception, EfficientNet, Inception, DenseNet, ResNet, and the deeper version (VGG19, ResNet152V2,...). The results demonstrated that all the pre-trained CNNs models performed very well, with Acc > 86.4%. Moreover, we found that ResNet152V2-based model showed the best performances for all metrics: Acc 95.5%, AUC 93.2%, Sen 94.4%, Pre 96.4%, and F1-score 95.2%. To interpret the model, we used CAM(Classification Activation Model).

Overall, in the above research, several studies have used XAI to apply in the interpretation of CAD. However, the explanation just has intuitive and is not deep, as well as the application of different explanatory methods to compare the effectiveness of the research, making the explanatory models unreliable. Therefore, further studies and experiments are needed in this area. With the traditional metrics insertion and deletion methods[20], the weights fluctuate from 0 to 1. This is only effective for the multi-label model because when deleting or inserting features, the probability will be divided among other labels. As for binary labels, when deleting or inserting, the probability of the label is not evenly divided, but there will be a preference for a label. Therefore, we propose the metrics deletion and insertion method using the threshold limit according to the area of the explanatory region.

III. MATERIALS AND METHODS

A. *CAD Dataset*

This paper is a follow-up to our previous work [3]. So we use the same data set to study for this paper, the data includes SPECT MPI polar maps from 218 patients collected in the Department of Nuclear Medicine of 108 Hospital, Hanoi, Vietnam. This dataset was obtained after a processing procedure and consultation with many technicians and doctors.

B. *Proposed methods*

As described in Fig.1, the methodological flow includes the following four parts: (i) Loading the dataset and preprocessing; (ii) Training model with Transfer Learning; (iii) Explaining the model with LIME, Grad-Cam, and RISE; (iv) Visualize, Metrics Insertion and Deletion. The following is a detailed explanation for each part:

(i) Loading dataset and preprocessing

The dataset has been preprocessed and divided into a training set consisting of 174 train images and 44 test images.

(ii) Traning model with Transfer Learning

This paper uses the Resnet152V2-based model, which showed the best performance in the previous research [3]. The results obtained on the test are Acc 95.5%, AUC 93.2%, Sen 94.4%, Pre 96.4%, and F1-score 95.2%. We use XAI methods to interpret this model and evaluate it against metrics.

(iii) Explain Model with LIME, Grad-Cam and RISE

a) LIME

LIME [10] is a post hoc method by making small increments instead of explaining the whole thing, we bring the facts to the local and show which part of the data has the most influence on our model predictions.

Algorithm 1 Sparse Linear Explanations using LIME

Require: Classifier f, Number of samples N
Require: Instance x, and its interpretable version x'
Require: Similarity kernel π_x Length of explanation K
Z ← {}
For i ∍ {1,2,3,…N} **do**
Z'I ←sample_around(x')
Z← Z U (z'_I,f(z_i), π(z_i))
end for
w← K-Lasso(Z,K)
return w

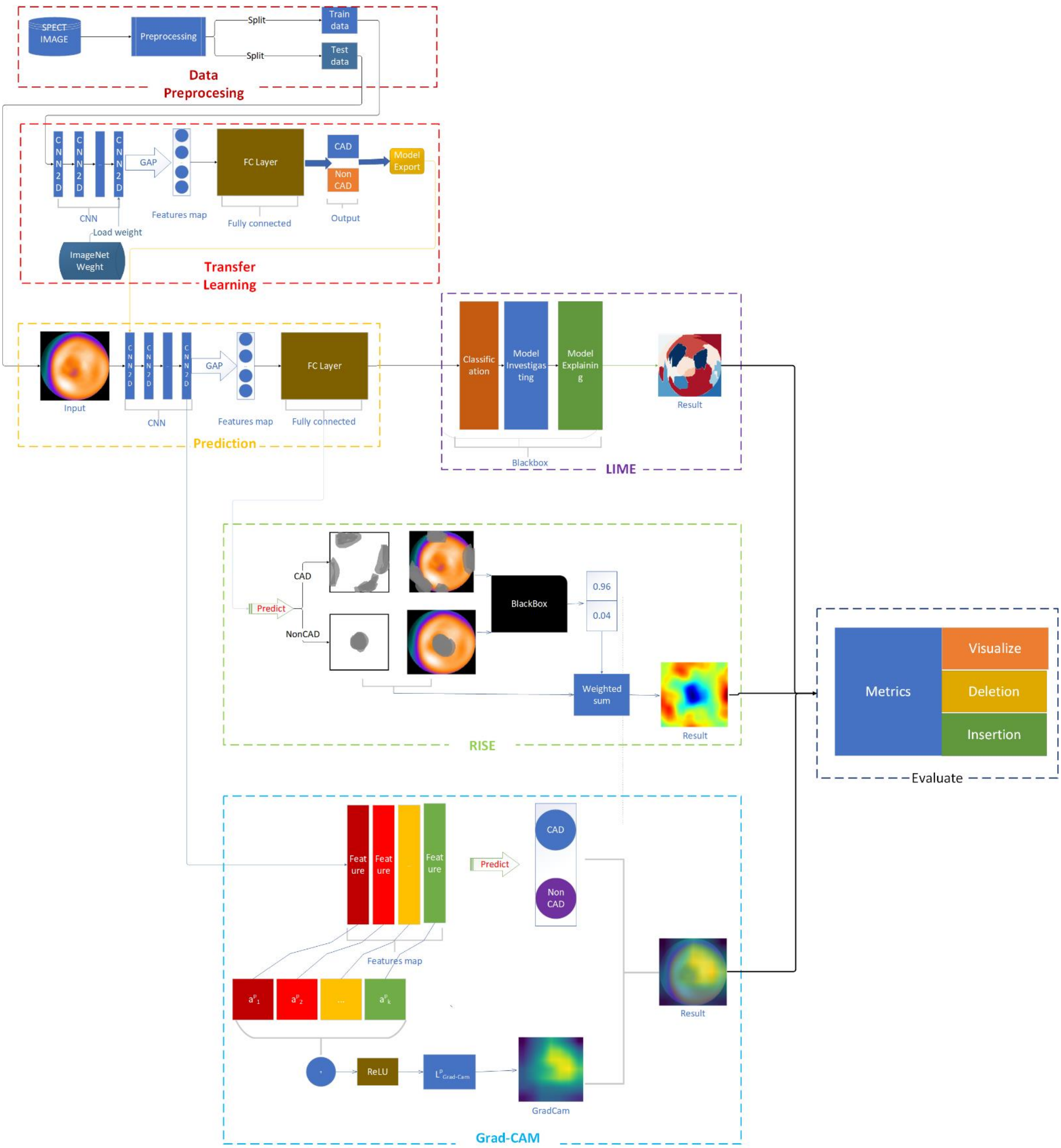

Fig. 1. Detailed explainability approach using LIME, Grad-Cam, and RISE for CAD Classification

b) GradCAM

GradCam [15] is a generalized instance of the CAM used in [10], which can be used with any network architecture. The idea of Grad-CAM is similar to CAM, and we use spatial information contained in convolution layers to understand which regions are essential for a particular class. Grad-CAM can be used on any feature maps from any layer in the network, but to demonstrate the efficiency and explanatory power of the model, the Grad-CAM paper uses the last convolution layer because the last convolution layer usually contains the most information. A^k in (1) is the Features map

$$L^{p}_{\text{Grad-CAM}} = ReLU(\Sigma_k a^p_k A^k) \quad (1)$$

c) RISE

RISE[14], Explains by measuring the importance of an image region as obscuring or 'contaminating' that region and observing how this affects black-box decisions. This is done by setting the pixel's intensity to zero, which opens the region or adds noise. RISE models by multiplying an image by [0,1] value masked. From there, the area affected by the decision of the model will be determined.

$$S_{I,f}(\lambda) \overset{\text{MC}}{\approx} \frac{1}{\mathbb{E}[M]\cdot N}\sum_{i=1}^{N} f(I\odot M_i)\cdot M_i(\lambda). \quad (2)$$

(iv) Visualize, Metrics Insertion and Deletion

In this section, we present the interpretations of LIME, GradCAM, RISE and evaluate the effectiveness of the interpretation.

To evaluate the effect of interpretation, we propose two auto-evaluation metrics: deletion, insertion, researched by [20]. We change the threshold of the algorithm. Instead of fixed value from 0 to 1, we use early stopping according to the weighted ratio of the interpreted positive area.

a) Deletion

Behind the deletion, the metric is the removal of the cause that will force the underlying model to change its decision. Specifically, this metric measures the decrease in the probability of the prediction class as more and more essential pixels are removed, where the importance is derived from the importance map. A sharp drop and, therefore, a low area along the probability line (as a function of the percentage of pixels removed) means a good explanation.

Algorithm 2 Traditional Deletion to compute deletion score

Procedure Deletion
  Input: black box *f*, image *I*, importance map *S*, number of pixels *N* removed per step
  Output: deletion score *d*
  $n \leftarrow 0$
  $h_n \leftarrow f(I)$
  while *I* has non-zero pixels **do**
    According to *S* ,set next *N* pixels in *I* to 0
    $n \leftarrow n+1$
    $h_n \leftarrow f(I)$
  $d \leftarrow$ AreaUnderCurve(h_i *vs* i/n, $\forall I = 0,\ldots n$)
  return *d*

Algorithm 3 Proposed Deletion to compute deletion score

Procedure Deletion
  Input: black box *f*, image *I*, importance map *S*, number of pixels *N* removed per step
  Output: deletion score *d*
  $n \leftarrow 0$
  threshold ← *Area*(S)/*Area(I)*
  $h_n \leftarrow f(I)$
  while *I* has *threshold* pixels **do**
    According to *S* ,set next *N* pixels in *I* to *threshold*
    $n \leftarrow n+1$
    $h_n \leftarrow f(I)$
  $d \leftarrow$ AreaUnderCurve(h_i *vs* i/n, $\forall I = 0,\ldots n$)
  return *d*

b) Insertion

Insertion metric, on the other hand, takes a complementary approach. It measures the increase in probability as more and more pixels are introduced, with a higher AUC of a better explanation.

Algorithm 4 Traditional Insertion to compute deletion score

Procedure Insertion
  Input: black box *f*, image *I*, importance map *S*, number of pixels *N* removed per step
  Output: insertion score *d*
  $n \leftarrow 0$
  $I' \leftarrow Blur(I)$
  $h_n \leftarrow f(I)$
  while $I \neq I'$ **do**
    According to *S* ,set next *N* pixels in *I'* to corresponding pixels in *I*
    $n \leftarrow n+1$
    $h_n \leftarrow f(I)$
  $d \leftarrow$ AreaUnderCurve(h_i *vs* i/n, $\forall I = 0,\ldots n$)
  return *d*

Algorithm 5 Proposed Insertion to compute deletion score

Procedure Insertion
  Input: black box *f*, image *I*, importance map *S*, number of pixels *N* removed per step
  Output: insertion score *d*
  $n \leftarrow 0$
  $I' \leftarrow Blur(I)$
  $h_n \leftarrow f(I)$
  threshold ← *Area*(S)/*Area(I)*
  threshold
  while $I \neq I'$**threshold* **do**
    According to *S* ,set next *N* pixels in *I'*threshold* to corresponding pixels in *I*
    $n \leftarrow n+1$
    $h_n \leftarrow f(I)$
  $d \leftarrow$ AreaUnderCurve(h_i *vs* i/n, $\forall I = 0,\ldots n$)
  return *d*

IV. RESULT AND DISSCUSION

A. *Results*

1) *Explanations*

a) LIME

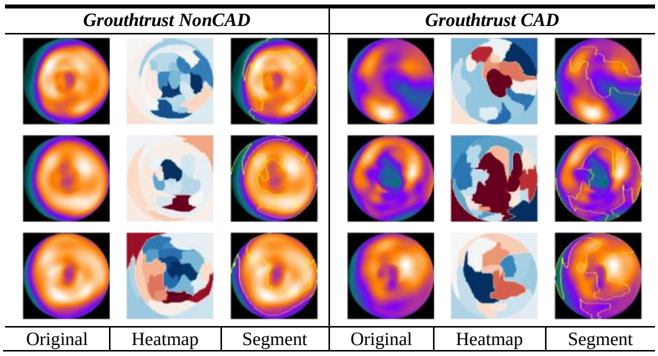

Fig. 2. Visualization LIME technique

Fig. 2 Presenting explanations for LIME, "Original" original SPECT MPI images; "Heatmap" describe the areas of explanation and influence on the model prediction, the green area is positive, the dump area is negative; "Segment" visualizes results generated by LIME.

In the interpretation for NonCAD, looking at the heatmap and segment, we can see that the blue area (for heatmap) or circled (for the segment) is the explanation area for NonCAD prediction, which is relatively accurate when describing the bright color SPECT image area and for CAD describing into the dark SPECT image area (the damaged area). The LIME Explanation method is very effective for proposed models.

b) GradCAM

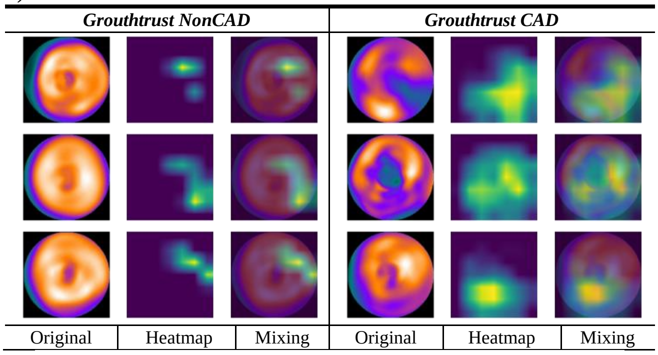

Fig. 3. Visualization GradCam technique. “Original” original SPECT MPI images; “Heatmap” explanation of GradCam, “Mixing” visualized result generated by Grad-CAM

c) RISE

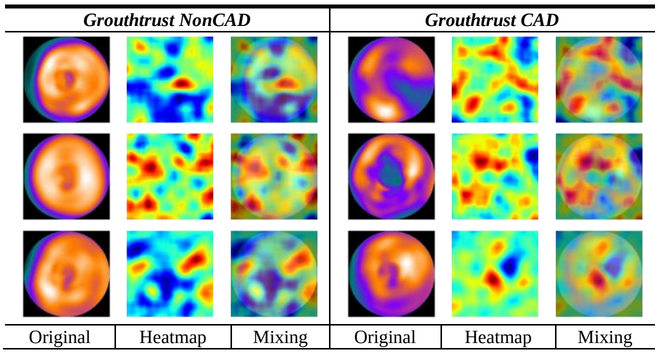

Fig. 4. Visualization RISE technique

Fig. 4 presents the explanations for RISE. “Original” original SPECT MPI images; “Heatmap” explanation of GradCam, “Mixing” visualized result generated by RISE. The RISE Explanation method is very effective for proposed models.

The interpretation for explained is highlighted in red, NonCAD is very standard when looking at the bright areas of the spectrum, and for CAD, the interpreted area is relatively large and is the dark (damaged area) of the SPECT image.

2) Metric

We visualization below to compare deletions and insertions for two methods: traditional [20] and proposed. The vertical y-axis describes the prediction probability of the label to be indexed, and the horizontal x-axis describes the area of the area to be deleted and inserted, with traditional from 0 to 1; proposed from 0 to threshold, All scaled to 0-1.

a) LIME

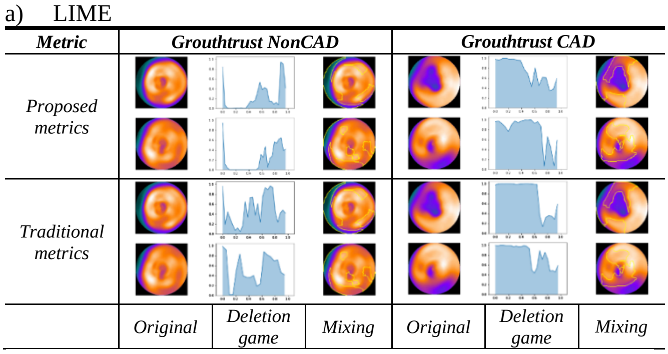

Fig. 5. Visualization LIME Deletion Technique; "Original" original SPECT MPI images;”Deletion game” visualization for prediction probability using deletion algorithm; “Segment” visualized results generated by LIME

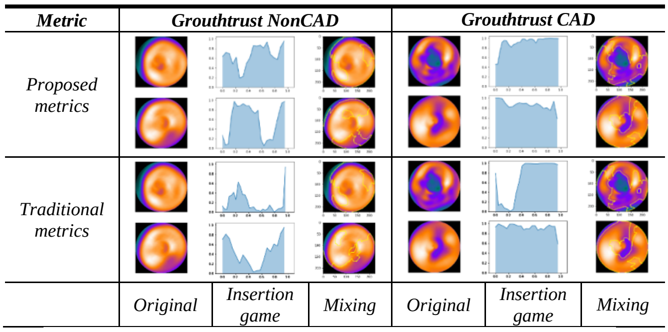

Fig. 6. Visualization LIME Insertion Technique; "Original" original SPECT MPI images; "Insertion game" visualization for prediction probability using insertion algorithm; "Segment" visualized results generated by LIME

b) GradCAM

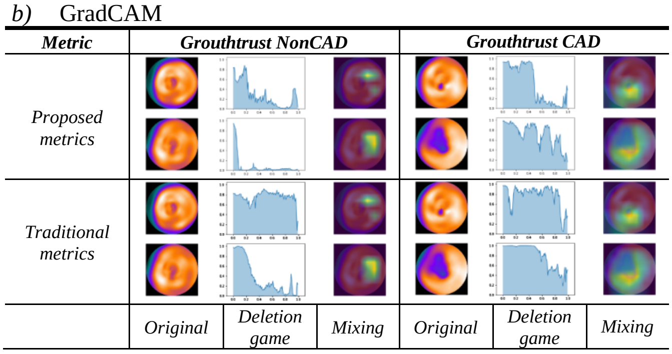

Fig. 7. Visualization GradCAM Deletion Technique; "Original" original SPECT MPI images; "Deletion game" visualization for prediction probability using insertion algorithm; "Mixing" visualized results generated by GradCAM

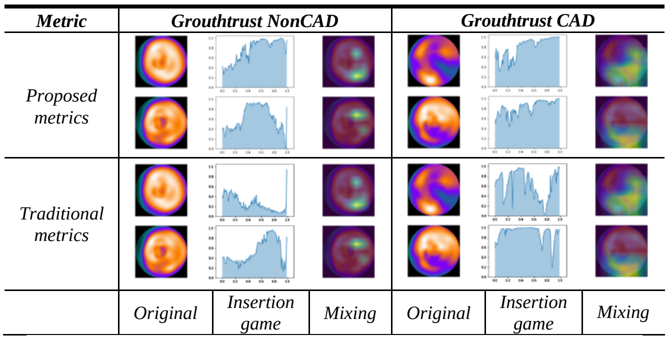

Fig. 8. Visualization GradCAM Insertion Technique; "Original" original SPECT MPI images; "Insertion game" visualization for prediction probability using insertion algorithm; "Mixing" visualized results generated by GradCAM

c) RISE

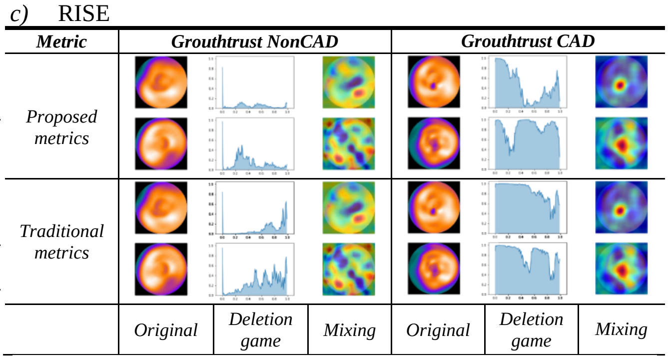

Fig. 9. Visualization RISE Deletion Technique; “Original” original SPECT MPI images; "Deletion game" visualization for prediction probability using insertion algorithm; "Mixing" visualized results generated by RISE

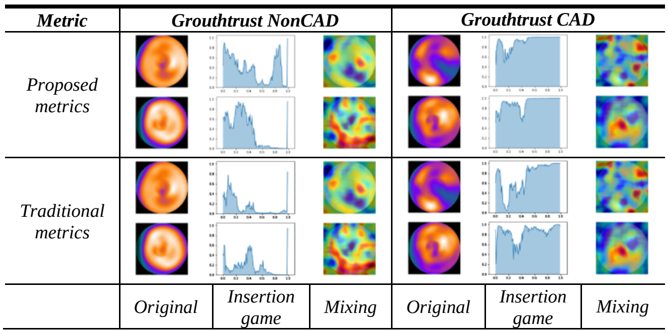

Fig. 10. Visualization RISE Insertion Technique; "Original" original SPECT MPI images; "Insertion game" visualization for prediction probability using insertion algorithm; "Mixing" visualized results generated by RISE

By applying the proposed algorithm, we have seen its effectiveness. In the deletion, instead of deleting all the features of the image, we only erase up to the threshold, and in the final result, the probability of the label scoring is 0.2-0.3 instead of 0.4 for NonCAD labels and 0.6 for CAD labels. In addition, the area of the Deletion game of the proposed algorithm is smaller than that of the traditional algorithm.

According to the traditional method, the image before being inserted has all the features deleted, and according to the proposed method, the image is deleted only in the area to be explained. The area of the Insertion game of the proposed algorithm is bigger than that of the traditional algorithm. The details are evaluated below.

Table. 1. Evaluates the average of insertions and deletions of three XAIs, comparing the performance of traditional and proposed metrics on the test dataset, including 44 images.

XAI	**Deletion**		**Insertion**	
LIME	*Traditional*	*Proposed*	*Traditional*	*Proposed*
	0.55	**0.51**	0.53	**0.58**
GradCAM	*Traditional*	*Proposed*	*Traditional*	*Proposed*
	0.52	**0.36**	0.61	**0.72**
RISE	*Traditional*	*Proposed*	*Traditional*	*Proposed*
	0.41	**0.38**	0.57	**0.63**

B. Discussions

In this study, we have developed three explanatory methods for the Resnet152V2-base model from our previous research [3]. As can be seen, CNNs do not provide transparency and interpretability in their decisions, which is an essential obstacle to their full integration into medical image analysis. Therefore, doctors cannot rely on the predictions provided. We have implemented LIME, GradCAM, and RISE techniques to solve it, generating heatmaps for interpretation. And the methods we used to give reasonably well explainable results are presented in SPECT MPI images, dark areas are CAD areas, and light and light areas are NonCAD areas.

In addition, we also implement two more techniques to improve Delete and Insert by changing the threshold to interpret the three methods LIME, GradCAM and RISE. It has overcome the disadvantages of traditional deletion and insertion methods when used for binary label classification interpretation.

In summary, all three interpretation methods for export can interpret images for CAD using SPECT MPI images. Overall, the present study constitutes an innovation in the interpretation of image classification models and evaluation of deletion and insertion metrics with promising results.

V. CONCLUSION

The proposed paper presents efforts known to develop an interpretable path to CAD diagnosis using SPECT MPI imaging and sophisticated modern interpretive and DL techniques. In addition to deploying a highly accurate model from research [3], it is necessary to address the ability to interpret images through visualization. For this study, the effectiveness of the LIME, GradCAM, and RISE interpretation tools was investigated and yielded promising results for the automatic and accurate diagnosis in nuclear cardiology. Doctors can use visualization techniques LIME, GradCAM, and RISE to compare and make effective and confident decisions, taking advantage of the visual explanations provided. . Therefore, LIME, GradCAM, and RISE methods have been proven to be effective tools in providing explanations for CNN-based decisions in SPECT MPI images.

In conclusion, this study contributes to the effective diagnosis of coronary artery disease. Thus it will promote confidence in using an interpretable artificial intelligence model for diagnosis in nuclear medicine.

REFERENCES

[1] Erito Marques de Souza Filho, Fernando de Amorim Fernandes, Christiane Wiefels, Lucas Nunes Dalbonio de Carvalho, Tadeu Francisco Dos Santos, Alair Augusto Sarmet M D Dos Santos, Evandro Tinoco Mesquita, Flávio Luiz Seixas, Benjamin J W Chow, Claudio Tinoco Mesquita, Ronaldo Altenburg Gismondi, "Machine Learning Algorithms to Distinguish Myocardial Perfusion SPECT Polar Maps," 2021 Nov, DOI: 10.3389/fcvm.2021.741667.

[2] Nikolaos I Papandrianos, Anna Feleki, Elpiniki I Papageorgiou, Chiara Martini, "Deep Learning-Based Automated Diagnosis for Coronary Artery Disease Using SPECT-MPI Images," 2022 July, DOI: 10.3390/jcm11133918

[3] P. N. Hai, N. C. Thanh, N. T. Trung, T. T. Kien, "Transfer Learning for Disease Diagnosis from Myocardial Perfusion SPECT Imaging," *Computers, Materials and Continua*, Vol.3. pp. 5925-5941, 2022 July, DOI:10.32604/cmc.2022.031027.

[4] Fagan, L. M.; Shortliffe, E. H.; Buchanan, B. G. (1980), "Computerbased medical decision making: from MYCIN to VM. Automedica," *Heuristic Programming Project*, Departments of Medicine and Computer Science Stanford University, Stanford, California.

[5] Alizadeh, Fatemeh (2021). "I Don't Know, Is AI Also Used in Airbags?: An Empirical Study of Folk Concepts and People's Expectations of Current and Future Artificial Intelligence". Icom. 20 (1): 3–17. doi:10.1515/icom-2021-0009. S2CID 233328352.

[6] Brown, John S.; Burton, R. R.; De Kleer, Johan, "Pedagogical, natural language, and knowledge engineering techniques," *SOPHIE I, II, and II. Intelligent Tutoring Systems. Academic Press*, Vol.4., pp.98-111, ISBN 0-12- 648680-8, 2016

[7] Bareiss, Ray; Porter, Bruce; Weir, Craig; Holte, Robert, Protos, "An Exemplar-Based Learning Apprentice. Machine Learning," *Morgan Kaufmann Publishers Inc*, Vol. 3., pp. 112–139, ISBN 1-55860-119-8, 2019.

[8] Bareiss, Ray, "Exemplar-Based Knowledge Acquisition: A Unified Approach to Concept Representation, Classification, and Learning. Perspectives in Artificial Intelligence," 2001.

[9] Tickle, A. B.; Andrews, R.; Golea, M.; Diederich, J. ,"The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural network," *IEEE Transactions on Neural Networks*, Vol.5. , No. 10-12, pp 1057-1068, 2018 Nov.

[10] Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, "Why Should I Trust You?" Explaining the Predictions of Any Classifier," *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, Vol.2., No.1-3, pp 97-101, 2016 June, DOI: 10.18653/v1/N16-3020.

[11] Scott M. Lundberg, SuIn Lee: "A Unified Approach to Interpreting Model Predictions," *School of Computer Science University of Washington Seattle, WA 98105*, 2017.

[12] Pang Wei Koh, Percy Liang, “Understanding Black-box Predictions via Influence Functions,” 2017.
[13] Mukund Sundararajan, Ankur Taly, Qiqi Yan, “Axiomatic Attribution for Deep Networks,” 2017.
[14] Vitali Petsiuk, Abir Das, Kate Saenko, “RISE: Randomized Input Sampling for Explanation of Black-box Models,” 2018.
[15] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradientbased Localization,” 2019
[16] Nikolaos I. Papandrianos Anna Feleki Serafeim Moustakidis Elpiniki I. Papageorgiou Ioannis D. Apostolopoulos Dimitris J. Apostolopoulos, “An Explainable Classification Method of SPECT Myocardial Perfusion Images in Nuclear Cardiology Using Deep Learning and Grad-CAM,” 2021
[17] Liu, H.; Wu, J.; Miller, “Diagnostic Accuracy of Stress-Only Myocardial Perfusion SPECT Improved by Deep Learning,” *Eur. J. Nucl. Med. Mol. Imaging*, Vol.48., pp. 2793–2800, 2021.
[18] Otaki, Y.; Singh, A.; Kavanagh, P.; Miller, R.J.H.; Parekh, T.; Tamarappoo, B.K.; Sharir, T.; Einstein, A.J.; Fish, M.B.; Ruddy, T.D, “Clinical Deployment of Explainable Artificial Intelligence of SPECT for Diagnosis of Coronary Artery Disease,” *JACC Cardiovasc. Imaging 2021*, Vol.4. , No.3-5. , pp. 99-110, 2019
[19] Ruth C Fong and Andrea Vedaldi, “Interpretable Explanations of Black Boxes by Meaningful Perturbation,” *In IEEE International Conference on Computer Vision 2017 Oct*, Vol.3., pp.115-121, 2017, DOI:10.1109/ICCV.2017.371
[20] Laura Ruis, Mitchell Stern, Julia Proskurnia, William Cha: “Insertion-Deletion Transformer,” 2020 Jan.