

Classification of coronary artery disease severity based on SPECT MPI polarmap images and deep learning: A study on multi-vessel disease prediction

DIGITAL HEALTH
Volume 10: 1–13
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241288430
journals.sagepub.com/home/dhj



Jui-Jen Chen^{1,*}, Ting-Yi Su^{2,*}, Chien-Che Huang³, Ta-Hsin Yang⁴,
Yen-Hsiang Chang¹ and Henry Horng-Shing Lu^{2,5,6} 

Abstract

Background: Coronary artery disease (CAD) is a global health concern. Conventional single photon emission computed tomography myocardial perfusion imaging (SPECT MPI) is a noninvasive method for assessing the severity of CAD. However, it relies on manual classification by clinicians, which can lead to visual fatigue and potential errors. Deep learning techniques have displayed promising results in CAD diagnosis and prediction, providing efficient and accurate analysis of medical images.

Methods: In this study, we explore the application of deep learning methods for assessing the severity of CAD and identifying cases of multivessel disease (MVD). We utilized the EfficientNet-V2 model in combination with DeepSMOTE to evaluate CAD severity using SPECT MPI images.

Results: Utilizing a dataset consisting of 254 patients (176 with MVD and 78 with single-vessel disease [SVD]), our model achieved an accuracy rate of 84.31% and area under the receiver operating characteristic curve (AUC) value of 0.8714 in predicting cases of MVD. These results underline the promising potential of our approach in MVD prediction, offering valuable diagnostic insights and the prospect of reducing medical costs.

Conclusion: This study emphasizes the feasibility of employing deep learning techniques for predicting MVD based on SPECT MPI images. The integration of Efficient-Net-V2 and DeepSMOTE methods effectively evaluates CAD severity and distinguishes MVD from SVD. Our research presents a practical approach to the early prediction and diagnosis of MVD, ultimately leading to enhanced patient outcomes and reduced healthcare costs.

Keywords

Coronary artery disease, multivessel disease, deep learning, SPECT MPI, EfficientNet-V2, DeepSMOTE

Submission date: 19 March 2024; Acceptance date: 10 September 2024

¹Department of Nuclear Medicine, Chang Gung Memorial Hospital, Kaohsiung Medical Center, Chang Gung University College of Medicine, Kaohsiung

²Institute of Statistics, National Yang Ming Chiao Tung University, Hsinchu

³Institute of Data Science and Engineering, National Yang Ming Chiao Tung University, Hsinchu

⁴Institute of Applied Mathematics, National Yang Ming Chiao Tung University, Hsinchu

⁵School of Post-Baccalaureate Medicine, Kaohsiung Medical University, Kaohsiung

⁶Department of Statistics and Data Science, Cornell University, Ithaca, NY, USA

*Contributed equally to this work.

Corresponding authors:

Yen-Hsiang Chang, Department of Nuclear Medicine, Chang Gung Memorial Hospital, Kaohsiung Medical Center, Chang Gung University College of Medicine, Kaohsiung.
Email: changyh@cgmh.org.tw

Henry Horng-Shing Lu, School of Post-Baccalaureate Medicine, Kaohsiung Medical University, 100 Shih-Chuan 1st Road, Kaohsiung, 80708, Taiwan.
Email: henryhslu@kmu.edu.tw



Introduction

Coronary artery disease (CAD) stands as a prominent and pervasive global health concern that poses a threat to the well-being of individuals worldwide. The primary culprit behind CAD is atherosclerosis, a condition characterized by the accumulation of plaque in the arteries, which results in the narrowing or obstruction of blood flow to the heart muscle. Early detection and precise prediction of its progression are of paramount importance in reducing mortality rates and enhancing treatment outcomes. The diagnostic methods for CAD include electrocardiogram (ECG), echocardiography, coronary computed tomography angiography (CCTA), computed tomography (CT), and myocardial perfusion imaging (MPI). Among these, MPI is the method used by nuclear medicine departments to diagnose CAD. It is a noninvasive examination, unlike the widely recognized and highly effective coronary angiography, which requires catheter insertion. MPI assesses myocardial blood flow to identify the severity of arterial stenosis and determines whether a patient has multivessel disease (MVD) or single-vessel disease (SVD). As the treatment for MVD and SVD differs, accurately identifying whether one or multiple vessels are blocked is crucial. During the test, patients are exposed to radiation within a safe range, and the cost is higher compared to ECG and echocardiography. The patient's current physiological state may also influence the results, potentially leading to false positives or negatives. Additionally, the interpretation of images requires experienced physicians. Therefore, using artificial intelligence (AI) to assist in diagnosis can improve accuracy, reduce diagnosis time and costs, and can also reduce the eye fatigue of doctors when interpreting images for a long time. Conventional single photon emission computed tomography myocardial perfusion imaging (SPECT MPI) furnishes three-dimensional data pertaining to a patient's stress and rest conditions. This noninvasive and cost-effective method plays a pivotal role in assessing the severity of CAD, ultimately reducing unnecessary medical investigations and aiding clinicians in delivering precise medical care.

When interpreting SPECT MPI images, clinicians must manually annotate and categorize each case, and the accuracy of the diagnosis largely hinges on the clinician's relevant experience. This process can be laborious and may readily lead to visual fatigue. Consequently, there arises a need for a system capable of assisting clinicians in the classification process, saving them valuable time, and curbing medical expenses. Furthermore, such a system can provide supplemental diagnostic information for less-experienced healthcare professionals.

Significant research endeavors have been devoted to exploring the application of deep learning in advancing computer-aided diagnostic systems. Deep learning, a potent machine learning technique, has garnered substantial

attention in the prediction and diagnosis of CAD in recent years. Deep learning models have showcased remarkable performance in a variety of medical image analysis tasks, including the identification of abnormalities in ECGs and SPECT MPI. Convolutional neural network (CNN) models, in particular, have demonstrated their capacity to yield rapid diagnostic results and alleviate clinician fatigue in the realm of visual diagnostics, as perceived by healthcare professionals.^{1,2}

Most of the past research related to detecting MVD has been based on calculating the systolic and diastolic volumes of blood vessels, without paying attention to imaging feature extraction.³ There are also methods to detect cardiovascular damage based on ECG, that can be combined with deep learning self-supervised transformer models.⁴ In addition, cadmium-zinc-telluride single photon emission computed tomography (CZT SPECT) images are combined with deep learning CNN models as an auxiliary diagnosis for myocardial ischemia.⁵ None of them use deep learning technology to automatically diagnose MVD through CZT SPECT images.

Garavand et al. conducted two studies on the application of machine learning algorithms in the diagnosis of CAD.^{6,7} The first study involved comparing seven commonly used machine learning classification algorithms, while the second explored the use of various machine learning and deep learning techniques in CAD diagnosis. In the comparison of the seven machine learning algorithms, performance metrics and area under the receiver operating characteristic curve (AUC-ROC) were used to evaluate each model's efficiency. The study found that random forest (RF) and support vector machine (SVM) algorithms had the highest efficiency. These algorithms, when combined with clinical data and cardiac imaging data, can play a significant role in predicting CAD. The broader study on machine learning techniques highlighted the application of CNNs in deep learning. CNN-based models can automatically perform feature selection and can be implemented on mobile devices or servers, allowing for real-time testing with simple equipment, and providing wide accessibility to clinical experts. The literature indicates that deep learning offers substantial assistance in the diagnosis of CAD.

This study's primary objective is to investigate the application of deep learning techniques for evaluating the severity of CAD in patients and identifying the presence of MVD. Specifically, our focus lies in assessing CAD severity based on SPECT MPI images. To accomplish this, we employed the EfficientNet-V2 model and DeepSMOTE methodology for classifying CAD impairment into MVD and SVD.

During the model training process, we encountered the challenge of sparse medical imaging data. To address this issue and enable the model to learn more lesion features, we employed DeepSMOTE, a synthetic oversampling technique that generates new samples by interpolating between existing ones. What sets DeepSMOTE apart from synthetic

minority oversampling technique (SMOTE) is its incorporation of the variational autoencoder technique from deep learning, resulting in generated images that closely resemble authentic clinical images. To validate the effectiveness of our approach, we conducted experiments using a dataset comprising 254 patients, with 176 diagnosed with MVD and 78 with SVD. Our model achieved an accuracy rate of 84.31% and an AUC value of 0.8714, underscoring its potential value in MVD prediction.

In summary, this study aimed to explore the feasibility of leveraging deep learning methods for MVD prediction and validated the approach of EfficientNet-V2 + DeepSMOTE through experiments on a dataset consisting of 254 patients. Of these, 176 were diagnosed with MVD and 78 with SVD. Our research introduces a novel and practical approach for the early prediction and diagnosis of MVD, with the potential to improve patient outcomes and reduce healthcare costs.

Data introduction

The participants of this study were patients with CAD, and data were collected from September 2019 to October 2021. The image source is retrospective data, all of which was provided by the Department of Nuclear Medicine at Kaohsiung Chang Gung Memorial Hospital. A total of 254 patients participated in this study, with 176 of them diagnosed with MVD, while the remaining 78 had SVD. The average age of the patients was approximately 68.28 years. The patients must be clinically diagnosed with CAD before MVD testing can be performed, so the images are from CAD patients. For the experiment, we selected 254 patients who met the following criteria: (1) They had undergone invasive coronary angiography (ICA) revealing significant coronary stenosis. (2) They exhibited a high pretest risk of CAD (>90%). Resting ECG displayed abnormalities. This group of patients served as the reference database for individuals with CAD.^{8–10} The treatment approach for CAD varies based on whether a patient has a single vessel blockage or multivessel blockage. Single-vessel blockages can often be managed with medication. In case the blockage exceeds 70% and involves two or more vessels, interventions such as percutaneous coronary intervention (PCI) with stent placement or coronary artery bypass surgery are typically required. It is important to note that for significant lesions in the left main coronary artery, PCI is not considered suitable.¹¹ Key characteristics of the patients, including age, weight, risk factors for CAD, and the number of diseased vessels, are provided in Table 1. The label of MVD was confirmed through invasive surgical procedures, specifically cardiac catheterization.

All patients used the stress-only data during the SPECT MPI, which is performed in two stages: the morning stress stage and the afternoon rest stage. This approach is adopted because when the stress stage SPECT MPI clearly indicates the absence of CAD, the patient is spared from

Table 1. Patient descriptive statistics.

	MVD (<i>n</i> = 176)	SVD (<i>n</i> = 78)	<i>p</i>
Age	69.6 ± 10.7	70.5 ± 11.3	0.571
Gender (M:F)	150:26	64:14	0.521
Weight	72.2 ± 51.2	67.7 ± 13.1	0.184
Body mass index	25.8 ± 3.6	25.2 ± 3.4	0.248
Hypertension	140 (79.5%)	57 (73.0%)	0.254
Diabetes	107 (60.8%)	45 (59.2%)	0.642
Dyslipidemia	119 (67.6%)	52 (68.4%)	0.882
1-vessel disease	–	78 (100%)	
2-vessel disease	64 (39.8%)	–	
Left main and/or 3-vessel disease	112 (60.2%)	–	

undergoing the rest stage, thus reducing the burden on the patient.^{12–14} This approach is also recommended by the International Atomic Energy Agency Nuclear Cardiology Protocols Study (INCAPS).¹⁵ The image data were analyzed using medical software—QPS and QGS software to generate unmasked SPECT MPI polar map images and masked SPECT MPI polar map images, respectively. QPS and QGS software offer clinicians a comprehensive solution for the analysis of SPECT MPI images. This combination of myocardial perfusion and left ventricular function assessment enhances the standardization and reproducibility of SPECT MPI interpretation. These software tools offer quantitative measurements and standardized reporting, which in turn improve interobserver agreement and facilitate longitudinal patient follow-up.¹⁴

Computing platform

Our study utilized the RTX 3060 GPU with 12GB VRAM and an Intel i7-13700 processor. We opted for Ubuntu 18.04.6 LTS as the stable operating system. For model construction and training, we employed Python 3.9.15 and PyTorch 1.13.0. GPU acceleration was enabled through CUDA 11.4, significantly enhancing training efficiency. With this hardware and software setup, we conducted experiments to assess the performance and effectiveness of the proposed method.

Methods

Statistical analysis

The imaging data of patients from the Department of Nuclear Medicine at Kaohsiung Chang Gung Memorial

Hospital, collected between January 2020 and December 2021, were used to assess whether there are significant differences in characteristics of patients with MVD and SVD. Statistical analyses were conducted using the t -test and chi-square test. For continuous variables, such as age, weight, and body mass index (BMI), the t -test was employed to compare the differences in means between the two groups. For categorical variables, such as gender, presence of hypertension, diabetes, and hyperlipidemia, the chi-square test was used for analysis and comparison. Both methods were applied under the assumption of independent samples. By calculating the p -value, we could determine if significant differences existed, which would allow us to further explore the impact of differences between patients with MVD and SVD on the study results. In this study, all p -values were greater than 0.05, indicating no significant differences. This suggests that the variables do not influence whether a patient falls into a particular type of CAD.

Image preprocessing

As the raw data used in the study contained a considerable amount of irrelevant information, we needed to crop the image. This cropping process involved the following steps:

1. *Rough cut*: In this initial step, we performed a rough cut to remove the excess or irrelevant parts of the image.
2. *Find the circle*: Subsequently, we located and identified the relevant circular area within the image.
3. *Precision cut*: With the circular region identified, we proceeded with a precision cut to refine the selection and isolate the specific area of interest.
4. *Cut the masked image*: Finally, we applied the necessary cropping to the masked image, ensuring that only the pertinent information was retained.

In this section, we will provide a detailed explanation for each of these steps in the method.

Rough cut. In the original image, only the SPECT MPI polar map image taken in the stress state is utilized. To extract this specific image, we initially perform a rough cut to define the target area. The cutting process adheres to the principle that the selected area must encompass the entire SPECT MPI polar map image captured during the stress state while excluding any irrelevant sections. The original image measures (1640×1080) pixels, and we perform the cut within the range of $x = [88, 568]$ and $y = [417, 892]$. This results in a final image size of (480×475) pixels. For a detailed depiction of the cutting range and content, please refer to Figure 1.

Find the circle. We employ the Hough method to accurately detect the boundaries of the polar map images, which take on a circular shape. The Hough method¹⁵ is an image processing technique renowned for its ability to identify

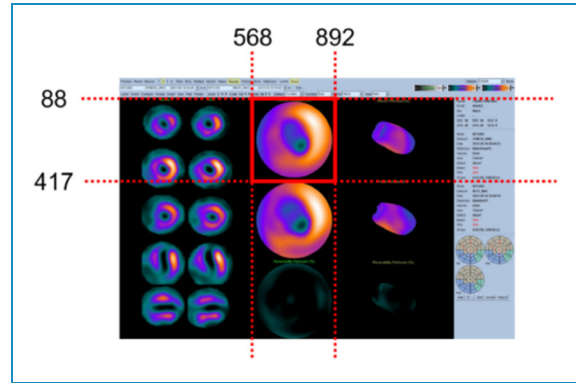


Figure 1. QPS system analysis screen-shot and the rough-cut image range.

circular shapes, making it valuable in numerous image processing applications. This method excels at tasks such as detecting round objects like coins or pupils, thanks to its high accuracy and robustness. In our case, we leverage the Hough method to detect circular shapes in the cardiac region, aiding us in the precise analysis of cardiac images.

In the Hough method, a circle is represented as a curve in a parameter space, where each point on the curve signifies a specific combination of a circle's center and radius. These combinations are determined by applying a suitable transformation formula to each edge pixel within the image.¹⁶ The circle detection process involves searching for peaks in the parameter space to identify the circles. Following circle detection, multiple candidate circles may emerge. However, in the context of our rough-cut result, there is typically only one circle representing our cardiac image range. To identify this circle, we select the largest one as our region of interest.

Precision cut. Following the extraction of the circular region, we proceed to mask the colors outside the circle, as the information beyond the circle is not directly relevant. To isolate the myocardial region in the rough-cut image, we use the smallest outer square of the circle as the boundary. The final outcome is visually depicted in Figure 2.

Cut the mask image. Due to the preprocessing conducted by the QPS system, the SPECT MPI polar map image taken in the stress state does not display the complete circular shape, rendering it challenging to accurately identify the correct circle area.

The error detection outcome is showcased in Figure 3(a). However, as the corresponding positions on the original masked image and the unmasked polar map image remain consistent, we can utilize the circular area from the unmasked polar map image to define the cropping boundaries for the masked image. The corrected range of the masked image is illustrated in Figure 3(b).

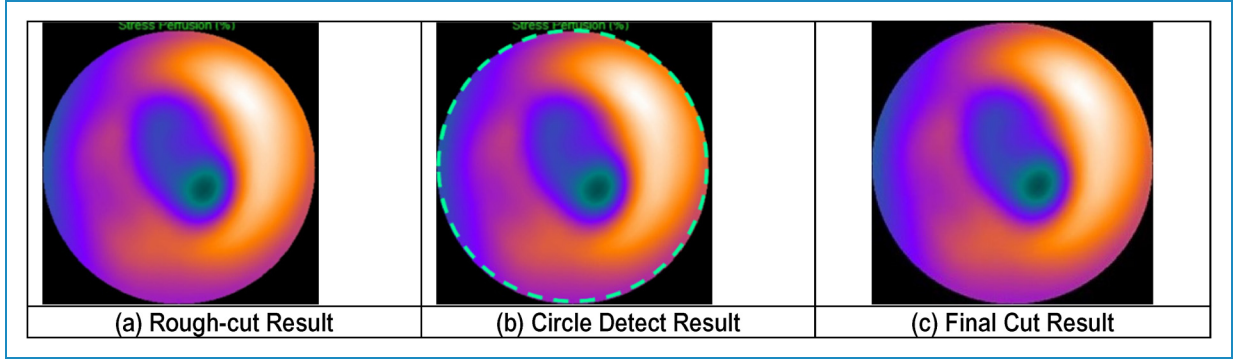


Figure 2. The result of image cut on each step. (a) Rough-cut result. (b) Circle detect result. (c) Final cut result.

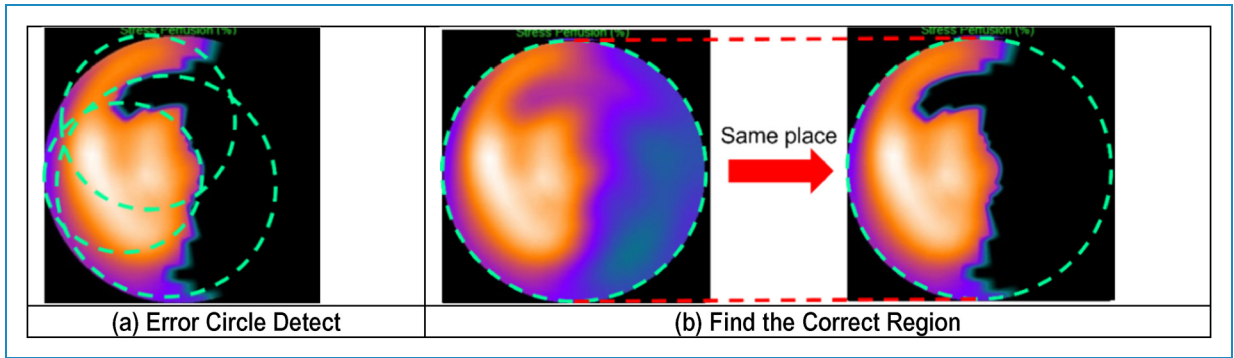


Figure 3. Cut the mask image. (a) Error circle detect. (b) Find the correct region.

Synthetic minority oversampling technique

Collecting data for deep learning is a formidable challenge, primarily because deep learning demands a significant volume of training samples that must meet rigorous standards in terms of quality and diversity, encompassing all possible scenarios and variations. In practice, assembling such datasets is often an arduous and costly endeavor. It consumes substantial time and human resources to annotate and organize the data effectively. Moreover, the challenge of data collection is exacerbated when dealing with class imbalance. In certain medical-related tasks, certain types of samples may be rare or challenging to discern. Patients typically undergo examination when they suspect they have a disease, and data concerning that disease are gathered. Consequently, there tends to be a scarcity of healthy data in the database, creating an imbalance in medical-related datasets. The issue of class imbalance can significantly impact the model's performance. It tends to prioritize the more frequently occurring class, which can lead to difficulties in accurately predicting the less frequent class.

Several studies have already proposed the use of synthetic data to address the issue of insufficient training datasets. In one study,¹⁷ liver tumor images generated with generative adversarial networks (GAN) were shown to effectively enhance the sensitivity of liver lesion

classification from 78.6% to 85.7% and improve specificity from 88.4% to 92.3%. The substantial increase in specificity from 88.4% to 92.3% underscores the effectiveness of the synthetic data method in enhancing the model's accuracy. In another study,¹⁷ GAN was employed to generate MPI polar map images, and a visual Turing test (VTT) was conducted to assess whether human observers could distinguish the synthetic images from actual ones. The initial prediction accuracy was 61.1%, with a standard deviation of 21.5%, which was not significantly higher than random guessing. Drawing from the insights of the above mentioned papers,^{18,19} numerous methods have been devised to tackle the data imbalance issue through the use of synthetic data. However, both of these studies hinged on GAN as the generative model—a method that proves challenging when the training dataset is limited.²⁰ In our approach, we leverage SMOTE as a solution to data generation and imbalance.²¹ SMOTE is a widely adopted oversampling method that excels at increasing the number of minority samples in a dataset by synthesizing new samples. This technique effectively balances the dataset, enabling the model to focus on learning the distinctive features of each class, rather than being biased by the number of instances in each class.

The fundamental concept behind SMOTE is the generation of new samples within the minority class. The

process involves selecting a minority sample and creating new instances by identifying the K-nearest neighbors of that sample. In Naive SMOTE, the process randomly selects the K-nearest neighbors for each chosen sample and conducts linear interpolation between the selected sample and each neighbor to produce a new sample. The quantity of new samples generated is determined by the desired level of over-sampling. SMOTE has proven to be highly effective in balancing class distributions and enhancing model performance across a range of applications, including credit fraud detection, medical diagnosis, and natural language processing. In the SMOTE algorithm, the quality of the different types of samples acquired significantly impacts the quality of the generated data. Therefore, various sampling methods like Borderline SMOTE, DeepSMOTE, and others have been developed to enhance the data generation process.

Borderline SMOTE proves particularly valuable when dealing with scenarios where the minority class is situated on the periphery of the majority class. In our dataset, it is challenging to visually confirm the proximity between the minority class (SVD) and the majority class (MVD). Hence, we employ Borderline SMOTE to generate these exceptions, thereby enhancing the model's classification capability. However, both Naive SMOTE and Borderline SMOTE are primarily suited for structured data. When processing unstructured data, such as video or audio, the quality of synthesis tends to be lower. This is because audio or video data require consideration of spatial information. When feeding data into Naive SMOTE and Borderline SMOTE, each feature, for instance, the three channels of image pixels, is treated as an independent point. This approach can lead to color artifacts during interpolation. To bridge this gap and create synthetic images that closely resemble actual ones while avoiding color anomalies, we introduce DeepSMOTE for data composition.

DeepSMOTE. DeepSMOTE is the extension of SMOTE, specifically tailored for deep learning models. In contrast to Naive SMOTE, which primarily relies on feature space

interpolation for synthetic sample generation, DeepSMOTE capitalizes on the feature representations learned by deep neural networks, resulting in the creation of more realistic synthetic samples. The DeepSMOTE process begins with the training of a deep neural network on the original imbalanced dataset, utilizing standard architectures like CNNs or recurrent neural networks. Subsequently, it employs feature space augmentation techniques to generate synthetic samples within the learned feature space. DeepSMOTE generates new samples by perturbing existing samples in the feature space, moving them along the direction of the gradient. This approach creates new samples that are more representative of the minority class. The input shape of each DeepSMOTE model layer is outlined in Figure 4 and consists of two main parts: an encoder and a decoder, both integral components of the DeepSMOTE architecture. The encoder is responsible for generating image features, which are then subjected to geometric SMOTE to produce new image features. After generating the new image feature, the decoder utilizes it to produce the synthetic image. Research studies^{22,23} have demonstrated the effectiveness of DeepSMOTE in significantly improving the performance of deep learning models when handling imbalanced datasets. Its applications span a wide array of domains, including medical image analysis, speech recognition, and natural language processing. One notable advantage of DeepSMOTE is its reliance on feature space interpolation, as opposed to feature interpolation. This difference results in more diverse and realistic synthetic samples, which, in turn, can mitigate the risk of overfitting. Nevertheless, it is essential to acknowledge certain limitations of DeepSMOTE.

In particular, it necessitates a pretrained deep learning model, which incurs training time and additional computational costs. Additionally, there is the potential for overfitting issues in the DeepSMOTE model, which may lead to inconsistencies in the quality of the generated images, particularly concerning the retention of disease features. In our experimental results, DeepSMOTE has exhibited superior

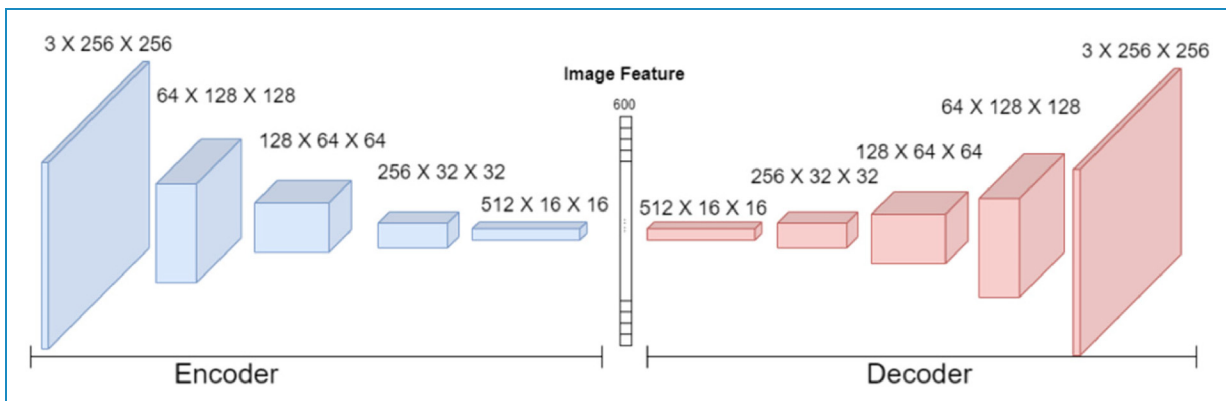


Figure 4. The input shape of each DeepSMOTE model layer.

predictive performance compared to other SMOTE methods and has effectively addressed the imbalance challenge encountered during the training of the deep learning model in this study.

Convolutional neural network models

The CNN models have recently achieved remarkable success in large-scale image and video recognition, thanks in part to the availability of extensive public image repositories like ImageNet.²⁴ The realm of 2D CNN boasts a plethora of pretrained models, some of which have gained considerable recognition, such as VGG-16, ResNet-18, Inceptionv3, and EfficientNet. When it comes to image recognition, the adoption of a pr-trained model offers several advantages. First, it obviates the need to invest substantial time in constructing a model's architecture. Second, the use of a pretrained model typically results in significantly higher accuracy compared to a custom-built CNN model. In this study, we opted to employ the architecture of EfficientNet-V2 as our model framework. The original EfficientNet-V2 was designed for classifying 1000 classes, but our task involves only two classes. Consequently, we retain the convolution layers and eliminate one of the fully connected layers, which originally had 1280 neurons. Furthermore, we employ the sigmoid function as the output activation function instead of the SoftMax function.

Performance metrics

Determining the quality of a model is not always straightforward, necessitating the use of specific indicators to assess its performance and serve as the basis for model selection. In the context of classification problems, the confusion matrix (as illustrated in Table 2) is a widely employed tool for evaluating performance, offering various metrics such as true negative, false negative, true positive, and false positive. These metrics, when combined and interpreted appropriately, help us select a primary performance indicator.

During the training phase, we identify the epoch with the highest validation metric as the final model. In binary classification problems, several common indicators are used to

gauge performance, including: (1) Accuracy: reflects the overall correctness of the model's predictions; (2) Sensitivity: also known as the true positive rate or recall, measures the ability to correctly identify positive instances; (3) Specificity: measures the ability to correctly identify negative instances; (4) Precision: reflects the model's ability to make accurate positive predictions; and (5) Balance accuracy: provides a balanced assessment of model performance across both positive and negative classes.

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

$$\text{Precision} = \frac{FP}{TP + FP} \quad (4)$$

$$\text{Balance accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (5)$$

Depending solely on metrics like accuracy, sensitivity, and specificity may offer a limited assessment of model performance, particularly when dealing with imbalanced data. The AUC addresses this limitation by considering sensitivity and specificity across various classification thresholds, resulting in a more comprehensive evaluation of the model's classification capability. AUC also provides a dependable means to compare the performance of different models, representing the probability of correct classification. Higher AUC values indicate superior classification ability.

In contrast, accuracy, sensitivity, and specificity offer insights at specific thresholds, making them less suitable for comprehensive model comparisons. Furthermore, AUC is valuable for assessing model stability. Unlike accuracy, sensitivity, and specificity, AUC exhibits minimal variation when confronted with minor dataset or prediction changes. In summary, incorporating AUC alongside other metrics enhances the assessment of a classification model's performance.

Results

Experiment data

In this study, we assess the performance of our model in addressing the MVD classification problem. Our dataset comprises a total of 254 records, with 176 falling under the category of MVD, while the remaining 78 are classified as SVD. Each data point includes information regarding the damage to the left anterior descending (LAD), left circumflex artery (LCX), and right coronary artery (RCA) blood

Table 2. Confusion matrix.

		Predicted	
		Negative (N)	Positive (P)
True	Negative (N)	True Negative (TN)	False Positive (FP)
	Positive (P)	False Negative (FN)	True Positive (TP)

vessels, with the number of affected blood vessels outlined in Table 3.

Synthetic data

We performed a series of experiments employing four different methods: no SMOTE, Naive SMOTE, Borderline SMOTE, and DeepSMOTE. The primary objective of these experiments was to assess the efficacy of these methods in enhancing the classification performance of our model. To address the imbalance in the dataset, we harnessed the power of Naive SMOTE, Borderline SMOTE, and DeepSMOTE to generate synthetic image data specifically for the minority class. The outcomes of these generations for both the unmasked and masked datasets are visually represented in Figures 5 and 6, respectively. Upon visual inspection, it becomes apparent that the

images generated through Naive SMOTE and Borderline SMOTE display peculiar blue or green hues. This anomaly can be attributed to the fact that both Naive SMOTE and Borderline SMOTE utilize interpolation techniques based on the original images. When the selected images encompass pixel regions of orange and purple colors, the interpolation process may inadvertently produce blue or green colorations. Furthermore, variations in the balance of orange and purple colors can give rise to the emergence of other distinctive color tones.

Model setting

In our research, we leveraged the pretrained EfficientNet-V2 model available in PyTorch to extract essential features from the input images. The PyTorch pretrained model, having been trained on the ImageNet dataset, typically includes a final classification layer that outputs predictions across 1000 different classes. In order to tailor this model to our specific binary classification problem, we made a modification by replacing the final classification layer with a single output neuron. Following the classification layer, the model produces a prediction score. It is important to note that this score is not a direct probability of having MVD. To transform the prediction score into a probability, we applied the sigmoid function. To determine the threshold for identifying MVD, we employed a

Table 3. Number of individual vessels damaged in SVD and MVD.

	LAD(+)	LCX(+)	RCA(+)
SVD	59	9	10
MVD	166	130	140
Total	225	139	150

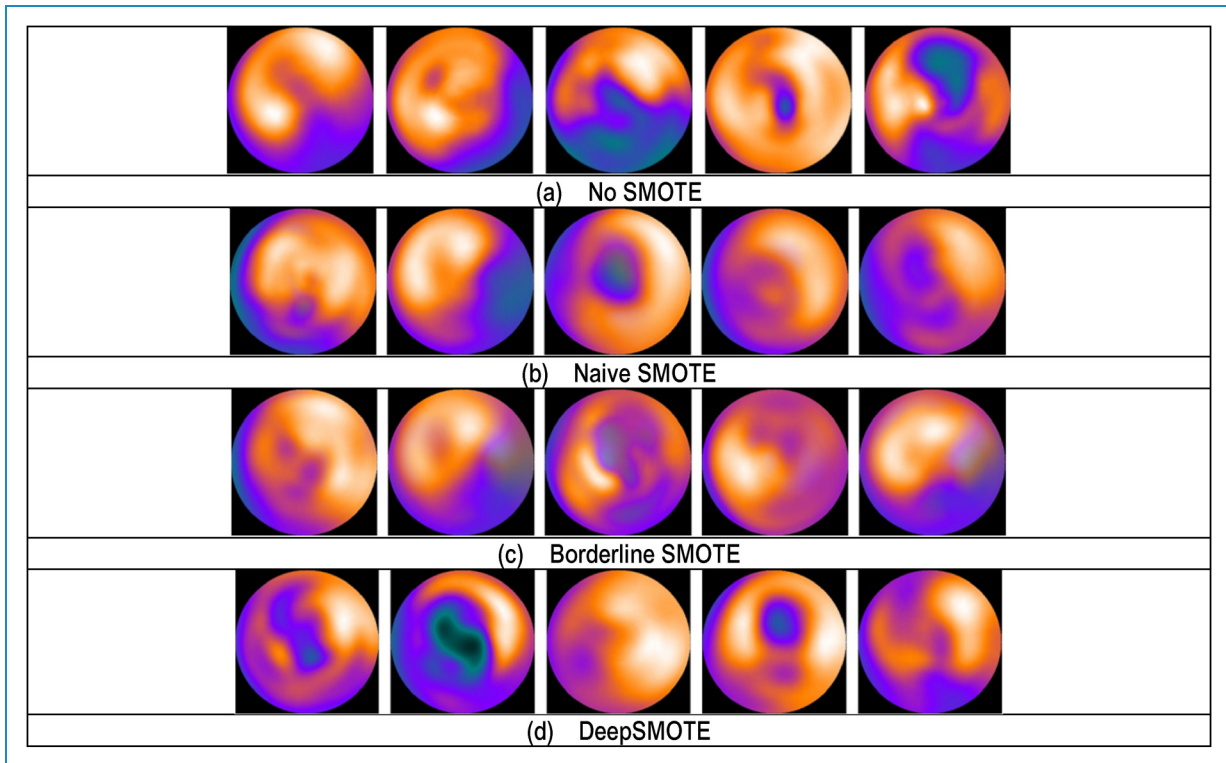


Figure 5. Unmasked image with four synthesis methods. (a) No SMOTE. (b) Naive SMOTE. (c) Borderline SMOTE. (d) DeepSMOTE.

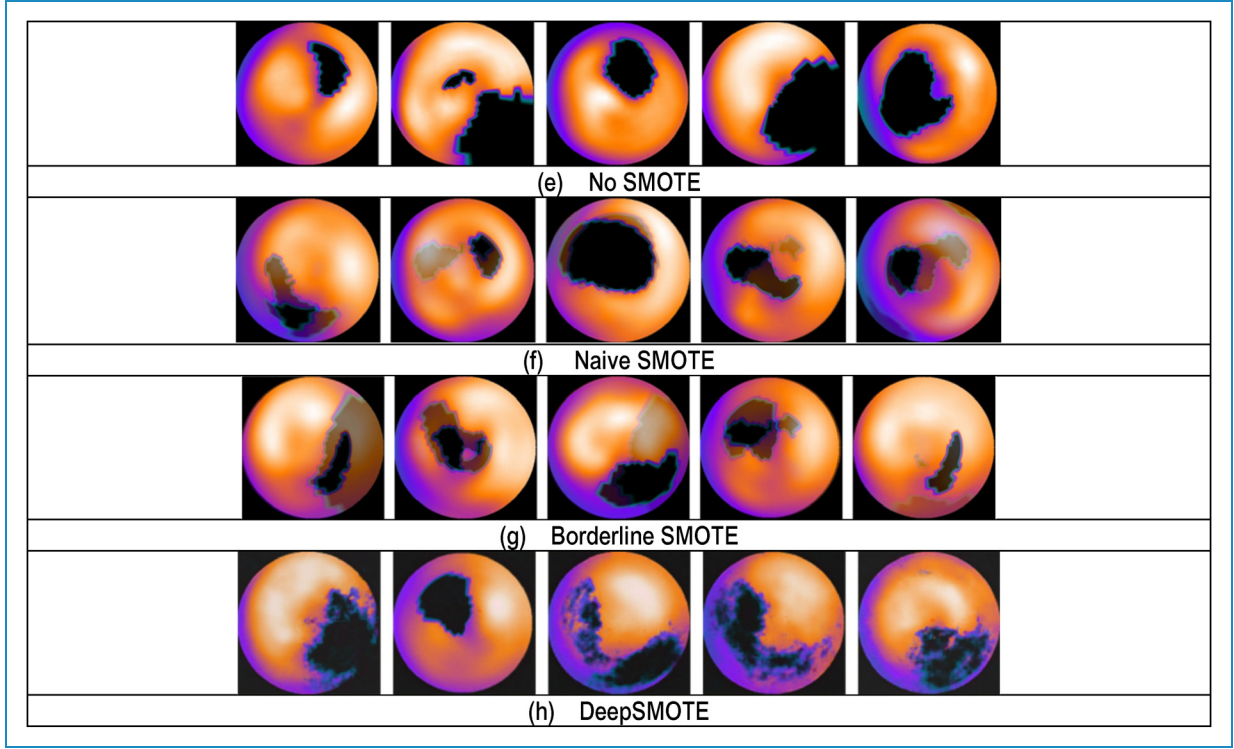


Figure 6. Masked Image with four synthesis methods. (a) No SMOTE. (b) Naive SMOTE. (c) Borderline SMOTE. (d) DeepSMOTE.

technique known as the Youden index. This index helps us select a threshold that maximizes the ability to distinguish between MVD and non MVD cases. Samples with probabilities exceeding this threshold are categorized as having MVD. During the training process, we employed the binary cross entropy (BCE) loss function, coupled with the Adam optimizer. To fine-tune the model's hyperparameters, specifically the learning rate and batch size, we conducted a grid search within the following ranges: learning rates from $[5 \times 10^{-4}, 5 \times 10^{-5}, 5 \times 10^{-6}]$ and batch sizes from (8,16,32). To address the challenge of data imbalance, we applied the weight balancing method within the BCE loss function, ensuring that the data remained balanced in scenarios where no SMOTE augmentation was used. We established a training epoch of 150 for each model, guaranteeing convergence, and saved the model with the highest validation AUC across all epochs.

Experiment results

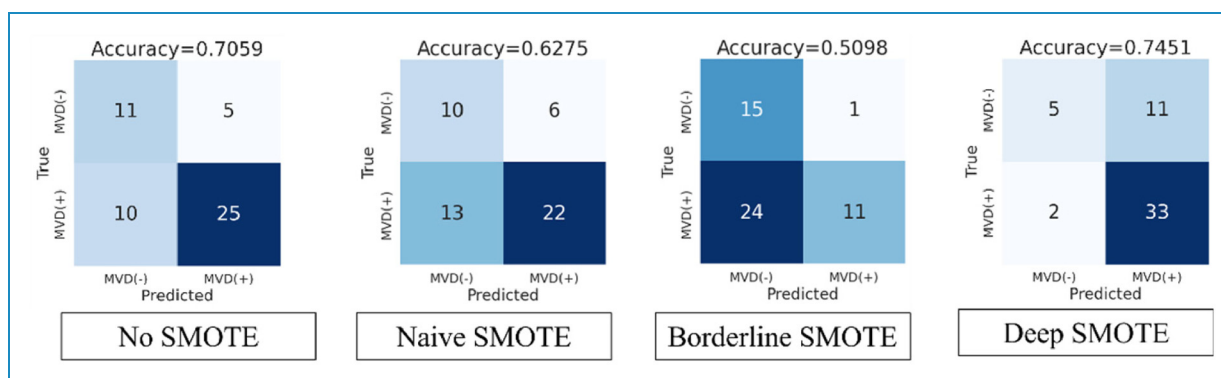
This study comprised two experiments designed to assess the impact of different oversampling methods and the utilization of masked images on the performance of deep learning models in image analysis. The experiment results are summarized in Table 4 and Figure 7, and are elaborated upon below, highlighting the key observations. We begin by examining the results for unmasked images. As depicted

in Figure 5, four oversampling methods were compared: No SMOTE, Naive SMOTE, Borderline SMOTE, and DeepSMOTE. The result of DeepSMOTE demonstrated the highest accuracy, achieving 74.51%, signifying its effectiveness in improving accuracy. Following closely is the No SMOTE method, registering an accuracy of 70.59%. In contrast, the Naive SMOTE and Borderline SMOTE methods yielded lower accuracy rates of 62.75% and 50.98%, respectively, indicating that their performance in this aspect was comparatively inferior. The result of DeepSMOTE excelled with a significantly higher sensitivity of 94.29%. This indicates that the method is exceptionally sensitive to identifying positive samples, particularly images featuring MVD characteristics. In contrast, Borderline SMOTE exhibited the lowest sensitivity at 31.42%. Notably, substantial discrepancies were also observed across the oversampling methods concerning specificity and Balance accuracy.

We now shift our attention to the results obtained when employing masked images, which offer more precise feature areas. As outlined in Table 5, the accuracy of all oversampling methods showed improvements when masked images were utilized. In particular, DeepSMOTE once again outperformed others, achieving an accuracy of 84.31%. This underscores the substantial positive impact of employing masked images on enhancing the model's accuracy. Moreover, the specificity of all oversampling methods exhibited improvement when masked images

Table 4. Unmasked result.

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	Balance accuracy (%)	F-Score (%)
No SMOTE	0.7196	70.59	71.43	68.75	70.09	76.92
Naive SMOTE	0.6875	62.75	62.85	62.50	62.67	69.84
Borderline SMOTE	0.6304	50.98	31.42	93.75	62.58	46.81
DeepSMOTE	0.7125	74.51	94.29	31.25	62.77	83.54

**Figure 7.** Confusion matrix of unmasked result.**Table 5.** Masked result.

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	Balance accuracy (%)	F-Score (%)
No SMOTE	0.8732	72.55	65.71	87.50	76.61	76.67
Naive SMOTE	0.7964	74.51	74.28	75.00	74.64	78.69
Borderline SMOTE	0.8107	72.55	71.42	75.00	73.21	78.13
DeepSMOTE	0.8714	84.31	85.71	81.25	83.48	88.24

were incorporated, leading to enhanced balance accuracy. However, it is noteworthy that the sensitivity of all over-sampling methods slightly decreased with masked images but remained relatively high. In summary, it can be concluded that the integration of masked images yields improvements in accuracy and specificity while potentially affecting sensitivity to a certain extent. Furthermore, DeepSMOTE closely approached the best-performing No SMOTE in terms of AUC and consistently delivered commendable accuracy. With the use of masked images, the balance accuracy of DeepSMOTE demonstrated significant superiority across all scenarios. This indicates that employing masked images alongside DeepSMOTE in MVD prediction tasks can effectively mitigate the impacts of data imbalance and enhance prediction performance (Figure 8).

Discussion

Over the past decade, statistics from the Ministry of Health and Welfare has consistently identified cardiovascular diseases as the second leading cause of death in the Taiwanese population. In the year 2022, Kaohsiung Chang Gung Hospital provided outpatient care for a total of 3497 patients with cardiovascular diseases.

Among these patients, a subset of 254 individuals underwent nuclear medicine examinations and received a diagnosis of CAD. The criterion for further assessment revolved around the presence or absence of MVD. To delve deeper into the factors contributing to CAD, statistical analysis was conducted, considering variables such as diabetes, lipid abnormalities, hypertension, smoking, and heart

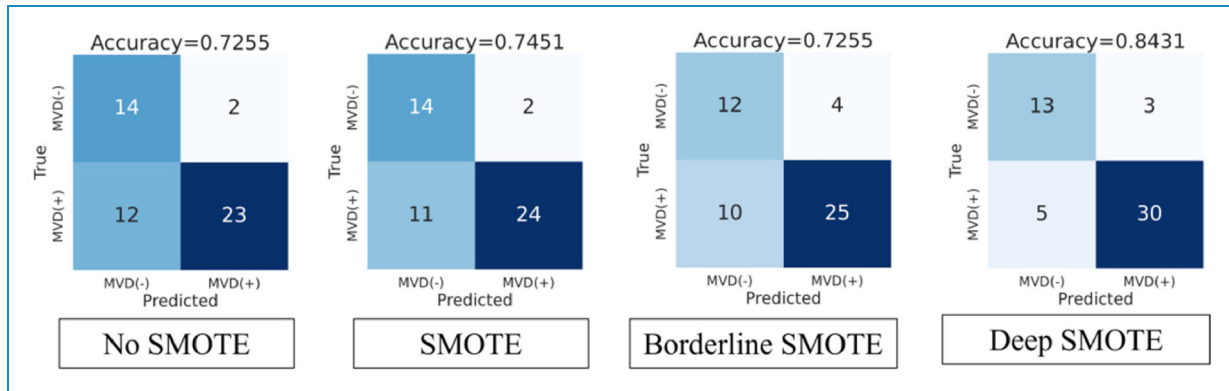


Figure 8. Confusion matrix of masked result.

failure. Although previous research has recognized various risk factors associated with CAD, the inconclusive significance observed in this study could potentially be attributed to limitations stemming from the sample size. Based on our experimental findings, we have identified distinct impacts associated with different oversampling methods on the models' performance in terms of classification ability. Notably, the Naive SMOTE and Borderline SMOTE methods were observed to be less effective in enhancing the models' predictive capabilities. In contrast, the DeepSMOTE method emerged as a standout performer, significantly boosting the model's accuracy in predicting MVD.

Moreover, when masked images were introduced, DeepSMOTE managed to further alleviate bias within the model's predictions and attain the highest balance accuracy. This observation bears substantial significance for our study. It underscores the fact that oversampling methods yield varying effects on model performance, and not all methods are equally proficient at enhancing classification ability. In this particular context, the DeepSMOTE method demonstrates distinct advantages in handling the MVD prediction task. It not only enhances accuracy but also excels in achieving superior balance accuracy by mitigating the adverse effects of data imbalance. In a related study, researchers achieved an AUC of 0.825 for MVD prediction using clinical information and logistic regression methods.²⁵ In comparison, our approach, which combines DeepSMOTE and masked image techniques, resulted in a relatively high performance with an AUC of 0.871. It is important to note that the DeepSMOTE method, due to its deep learning nature, demands more computational resources and time than other methods.

In summary, our research findings demonstrate the effectiveness of utilizing synthetically generated data through the SMOTE method in improving the classification performance of deep learning models on imbalanced datasets. By leveraging DeepSMOTE in conjunction with masked images, we have constructed a model that performs on par with the current best model in predicting MVD.

Although DeepSMOTE exhibits the most favorable performance among various SMOTE methods, it comes at the expense of increased computational resources and time requirements.

In the medical clinical use scenario, the DeepSMOTE model from this study first interprets stress images, providing doctors with red/yellow/green warning indicators in sequence: red for CAD, yellow for uninterpretable images, and green for no CAD. Doctors only need to focus on interpreting images with a red warning, which indicates myocardial defects, and compare these with rest images to write reports. This approach will significantly reduce the time doctors spend on image interpretation, thereby improving the quality of patient care.

In the future, we aim to prioritize cooperation with other Chang Gung Memorial Hall campuses to expand the sample size of the training set and enhance its predictive capabilities. The resulting model will be made available to other hospitals for external validation, thus promoting broader applicability and robustness.

Conclusion

Based on the existing research findings, it appears that despite the presence of numerous risk factors associated with CAD, these factors alone cannot reliably distinguish between patients with MVD and SVD. This lack of differentiation may be attributed to variations in sample sizes, potentially resulting in less pronounced significant relationships. Furthermore, it is plausible that the impact of these risk factors on disease progression may not yet be substantial enough. Therefore, the implementation of an interpretive service system proves highly valuable in the diagnosis of this condition. This system can offer valuable insights into disease diagnosis, even in situations where risk factors alone do not suffice to determine whether a patient has MVD or SVD.

In conclusion, the results of our study highlight the effectiveness of synthetic data generated through SMOTE-based

methods in enhancing the classification capabilities of deep learning models for unbalanced datasets. By employing the DeepSMOTE method in conjunction with masked images, we have developed a model that performs comparably to the current best model in predicting MVD. Among the various SMOTE methods, DeepSMOTE demonstrates the most favorable performance; however, it comes with higher computational resource and time requirements.

Acknowledgement: The authors thank Wan-Yi Tai for her valuable assistance and acknowledge the National Center for High-performance Computing for providing computing resources. Generative AI and AI-assisted technologies are declared in the writing process. During the preparation of this work, the authors used ChatGPT to revise the English language usage within the article. After using this tool/service, they reviewed and edited the content as needed and take full responsibility for the content of the publication.

Contributorship: All authors contributed to the research.


Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: This study was conducted by the guidelines and regulations provided by the Chang Gung Medical Foundation Institutional Review Board (protocol code 202101511B0C501 and approved September 15, 2021).

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work received funding from various sources, including the Chang Gung Memorial Hospital (Grant: CORPG8M0041), the National Science and Technology Council (Grants: 112-2634-F-A49-003-, 113-2118-M-A49-007-MY2, and 113-2923-M-A49-004-MY3), and the Higher Education Sprout Project of the National Yang Ming Chiao Tung University from the Ministry of Education, Taiwan.

Informed Consent: This research is a retrospective study. The medical images will be de-identified to maintain patient privacy. Furthermore, the research is categorized as minimal risk, meaning the potential risks to participants do not exceed those of nonparticipants. Waiving prior consent does not affect the rights of the study participants.

Guarantor: Yen-Hsiang Chang, Henry Horng-Shing Lu.

ORCID iD: Henry Horng-Shing Lu  <https://orcid.org/0000-0002-4392-3361>

References

1. Berkaya SK, Sivriköz IA and Gunal S. Classification models for SPECT myocardial perfusion imaging. *Comput Biol Med* 2020; 123: 103893.

2. Mostafapour S, Gholamiankhah F, Maroufpoor S, et al. Deep learning-guided attenuation correction in the image domain for myocardial perfusion SPECT imaging. *J Comput Des Eng* 2022; 9: 434–447.
3. Tanaka H, Chikamori T, Hida S, et al. The diagnostic utility of the Heston index in gated SPECT to detect multi-vessel coronary artery disease. *J Cardiol* 2008; 51: 42–49.
4. Moody JB, Rivière AP, Renaud JM, et al. Self-supervised deep representation learning of a foundation transformer model enabling efficient ECG-based assessment of cardiac and coronary function with limited labels. *MedRxiv* 2024; 2023: 10.25.23297552.
5. Su TY, Chen JJ, Chen WS, et al. Deep learning for myocardial ischemia auxiliary diagnosis using CZT SPECT myocardial perfusion imaging. *J Chin Med Assoc* 2023; 86: 122–130.
6. Garavand A, Salehnasab C, Behmanesh A, et al. Efficient model for coronary artery disease diagnosis: A comparative study of several machine learning algorithms. *J Healthc Eng* 2022; 2022: –9.
7. Garavand A, Behmanesh A, Aslani N, et al. Towards diagnostic aided systems in coronary artery disease detection: a comprehensive multiview survey of the state of the art. *Int J Intell Syst* 2023; 2023: 1–19.
8. Hendel RC, Berman DS, Carli D, et al. ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 appropriate use criteria for cardiac radionuclide imaging: a report of the American College of Cardiology foundation appropriate use criteria task force, the American society of nuclear cardiology, the American college of radiology, the American Heart Association, the American society of echocardiography, the society of cardiovascular computed tomography, the society for cardiovascular magnetic resonance, and the society of nuclear medicine endorsed by the American college of emergency physicians. *J Am Coll Cardiol* 2009; 53: 2201–2229.
9. Duvall WL, Slomka PJ, Gerlach JR, et al. High-efficiency SPECT MPI: comparison of automated quantification, visual interpretation, and coronary angiography. *J Nucl Cardiol* 2013; 20: 763–773.
10. Rubeaux M, Xu Y, Germano G, et al. Normal databases for the relative quantification of myocardial perfusion. *Curr Cardiovasc Imaging Rep* 2016; 9: 1–11.
11. HOSPITAL T-YG. Shared Decision-Making Assessment Form.
12. Gibson PB, Demus D, Noto R, et al. Low event rate for stress-only perfusion imaging in patients evaluated for chest pain. *J Am Coll Cardiol* 2002; 39: 999–1004.
13. Gutstein A, Bental T, Solodky A, et al. Prognosis of stress-only SPECT myocardial perfusion imaging with prone imaging. *J Nucl Cardiol* 2018; 25: 809–816.
14. Ueyama T, Takehana K, Maeba H, et al. Prognostic value of normal stress-only technetium-99m myocardial perfusion imaging protocol—comparison with standard stress-rest protocol. *Circ J* 2012; 76: 2386–2391.
15. Lindner O, Pascual TN, Mercuri M, et al. Nuclear cardiology practice and associated radiation doses in Europe: results of the IAEA Nuclear Cardiology Protocols Study (INCAPS) for the 27 European countries. *Eur J Nucl Med Mol Imaging* 2016; 43: 718–728.
16. Chen Y-L. *Nuclear myocardial perfusion imaging: comparison of image acquisition time with conventional SPECT and CZT detector based technology*. Hsinchu, Taiwan: National Yang Ming Chiao Tung University, 2012.

17. Stockman G and Shapiro LG. *Computer vision*. Hoboken, NJ: Prentice Hall PTR, 2001.
 18. Frid-Adar M, Diamant I, Klang E, et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018; 321: 321–331.
 19. Higaki A, Kawada Y, Hiasa G, et al. Using a visual turing test to evaluate the realism of generative adversarial network (GAN)-based synthesized myocardial perfusion images. *Cureus* 2022; 14: e30646.
 20. Liu H, Zhang W, Li B, et al. Improving GAN training via feature space shrinkage. *arXiv Preprint* 2023; arXiv: 230301559v2.
 21. Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16: 321–357.
 22. Murugan S, Venkatesan C, Sumithra MG, et al. DEMNET: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. *IEEE Access* 2021; 9: 90319–90329.
 23. Cini Oliveira M, de Araujo Eleuterio T, de Andrade Corra AB, et al. Factors associated with death in confirmed cases of COVID-19 in the state of Rio de Janeiro. *BMC Infect Dis* 2021; 21: 1–16.
 24. Deng J, Dong W, Socher R, et al. (eds) Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE; 2009.
 25. Kunita Y, Nakajima K, Nakata T, et al. Prediction of multi-vessel coronary artery disease and candidates for stress-only imaging using multivariable models with myocardial perfusion imaging. *Ann Nucl Med* 2022; 36: 674–683.
-