AI-based classification algorithms in SPECT myocardial perfusion imaging for cardiovascular diagnosis: a review

Nikolaos I. Papandrianos[a], Ioannis D. Apostolopoulos[b], Anna Feleki[a], Serafeim Moustakidis[a,c], Konstantinos Kokkinos[a] and Elpiniki I. Papageorgiou[a]

In the last few years, deep learning has made a breakthrough and established its position in machine learning classification problems in medical image analysis. Deep learning has recently displayed remarkable applicability in a range of different medical applications, as well as in nuclear cardiology. This paper implements a literature review protocol and reports the latest advances in artificial intelligence (AI)-based classification in SPECT myocardial perfusion imaging in heart disease diagnosis. The representative and most recent works are reported to demonstrate the use of AI and deep learning technologies in medical image analysis in nuclear cardiology for cardiovascular diagnosis. This review also analyses the primary outcomes of the presented research studies and suggests future directions focusing on the explainability of the deployed deep-learning systems in clinical practice.

Nuclear Medicine Communications XXX, **XXX**:XXXX–XXXX

Keywords: deep learning, classification, cardiovascular diseases, nuclear medicine, SPECT, artificial intelligence

[a]Department of Energy Systems, University of Thessaly, Larisa, Greece, [b]Department of Medical Physics, School of Medicine, University of Patras, Patras, Greece and [c]AIDEAS OÜ, Tallinn, Estonia

Correspondence to Nikolaos I. Papandrianos, Department of Energy Systems, University of Thessaly, Gaiopolis Campus, 41500 Larisa, Greece
Tel: +6936064614; e-mail:npapandrianos@uth.gr

Received 11 August 2022 **Accepted** 6 October 2022.

Introduction

Heart disease is the highest-rated cause of death globally, and more specifically, coronary artery disease (CAD) is the most prevalent type that occurs [1]. CAD is developed when the blood supply leading to the heart muscle is impassable due to a blocked coronary artery. The early detection of any signs of heart disease of any kind is crucial, thus providing an efficient treatment and easing a patient's life [2].

Single-photon emission computer tomography (SPECT) myocardial perfusion imaging (MPI) is an extensively used technique in analyzing heart images. This approach is a noninvasive diagnostic imaging procedure [3] that provides three-dimensional heart representation information in stress and rest mode. Furthermore, it is a cost-effective technique for CAD diagnosis that computes the severity of CAD, reduces unnecessary angiographies, and helps demonstrate the proper treatment. Expert clinical readers visually classify the labeling of each SPECT case; however, it is time-consuming and depends on the corresponding reader's related experience. Therefore, an autonomous system is needed to assist clinical experts with diagnosing and classifying the images, further achieving high accuracy and reducing healthcare costs [4,5]. Computer-aided systems and, more specifically, deep learning (DL) methodologies have rapidly been utilized in clinical diagnostics as they can reduce doctors' daily struggles [6–8]. Moreover, they have adopted magnificent approaches to avoid errors, which are being applied in diagnosing various diseases [9].

The recent development of innovative methods and techniques in image processing and analysis has been pushed by the exponential growth of data over the past few decades and the rapid expansion in the computing capability of modern computer systems. These applications range from simple and well-known artificial intelligence (AI) methods like artificial neural networks (ANNs) [10] and support vector machines (SVMs) [11] to more complicated and deep networks like convolutional neural networks (CNNs) [12] and generative adversarial networks (GANs) [13]. Particularly in the CAD domain, CNNs have been applied in numerous research studies regarding the prediction of abnormalities because they can learn patterns and perform perfectly with images [9].

The field of CAD classification is expected to be dominated by DL, especially CNNs. The vast research volume on CNNs is evidence of their efficacy and widespread usage. Various research communities are simultaneously creating these applications, yet the results of their dissemination are dispersed throughout an extensive array of conference proceedings and publications. The main contribution of the present work is to report related AI studies for CAD classification utilizing SPECT MPI images with respect to extracted metrics like accuracy, sensitivity, specificity and area under the curve (AUC). DL methodologies are being examined and compared against standard quantitative methods. The review study classifies the research papers according to their DL methods, their validation procedures in terms of data size and evaluation method, and their quantitative results. In the Discussion section, the study presents the most

DOI: 10.1097/MNM.0000000000001634

significant results and highlights major limitations that need future examination.

The present paper is organized into the following sections: Section 2 briefly describes the dominant machine learning (ML) and DL methods found in the literature. In Section 3, the review methodology, inclusion and exclusion criteria, as well as the literature sources are described. A detailed discussion is provided in section 4, highlighting the major findings and challenges, whereas the last section outlines the concluding remarks.

Machine and deep learning: main aspects

Machine learning

Machine learning (ML) is an AI system that aids algorithms to become more accurate by learning the extracted knowledge from data and successfully predicting the outcome without the need to program them fully. Learning is divided into unsupervised, supervised and semisupervised. Regarding unsupervised, the autonomous model creates patterns and forms groups using unlabelled data. In contrast, supervised learning uses labels along data, so the algorithm produces features through iterations of the dataset. Semisupervised learning is a combination of the above methods, where some data are labeled and some are not [6]. A key factor when employing ML is the split of the dataset into training, testing and validation. For example, the training dataset may consist of 85% of the total dataset, whereas the remaining 15% is used for testing. The validation dataset typically includes 20% of the training dataset. More specifically, the training dataset contains images and the corresponding output for pattern extraction while training the model. The validation dataset helps the model fine-tune its hyper-parameters after each epoch, which optimizes error. In contrast, the testing dataset extracts the final accuracy and loss in classifying unseen data [14].

Deep learning

DL is a subset of ML based on the application of neural networks (NNs). NNs derive their functionality from the biological synaptic systems. NNs include input, hidden, and output layers in the simplest form. More specifically, a deep NN is a network with multiple hidden layers. DL is the application of deep NNs [15].

Regarding the input layer, it can accept large amounts of data and transfer them to the hidden layer(s) for processing and extracting patterns, where each layer tries to produce more abstract features to achieve generability. The produced features of the hidden layers are transmitted to the output layer to extract the corresponding prediction [16]. A CNN is a type of deep NN that has demonstrated extreme efficiency in classifying images. It can encounter each pixel as a value and insert them into an array [15,16]. In general, CNN models design a hierarchical structure of extracting features. Their most significant advantage is that there is no need for predefined image features [16]. With respect to their extracted results, CNNs have become an integral tool in medical image analysis since they can retrieve hierarchical features by processing high-level from low-level features [17].

DL techniques enable people to (1) automatically learn from data, (2) detect underlying patterns and eventually (3) make efficient decisions. DL provides an advanced analytics tool for medical diagnosis with automatic feature learning and high-volume modeling capabilities. It uses a cascade of layers of nonlinear processing to learn the representations of data corresponding to different levels of abstraction. The hidden patterns underneath each other are then identified and predicted through end-to-end optimization. Thus, DL offers great potential to boost data-driven medical decision-making applications.

The primary ML and DL algorithms, as presented in the publications included in this study, are summarized in further detail in Table 1, which can be found below.

Materials and methods

The present review follows the process defined in the reporting items for systematic reviews and meta-analyses (PRISMA) guidelines [24]. Each step of the review process (literature search, databases, and study selection) was independently performed by two authors (N.P., A.F.). Discrepancies were resolved by a third author (E.P.).

Literature search

A structured literature search was conducted in the following databases: (a) MEDLINE (through PubMed) and (b) SCOPUS. Articles cited by the retrieved papers, as well as articles citing the retrieved papers (using Google Scholar), were also identified by conducting a supplementary manual search. The potential eligibility of the articles was initially decided based on their titles and abstracts. The authors investigated full texts to verify whether the initial qualifiers fulfilled the inclusion criteria. The structured search strategy per database is quoted below.

Databases

MEDLINE: ((Nuclear medicine) OR (SPECT)) AND (Coronary Artery Disease) AND ((deep learning) OR (convolutional neural networks) OR (artificial intelligence) OR (generative adversarial networks) OR (explainable) OR (machine learning)).

SCOPUS: TITLE-ABS-KEY ((‘coronary artery disease’ AND ‘SPECT’ AND (‘deep learning’ OR ‘artificial intelligence’ OR ‘convolutional neural networks’ OR ‘explainable’)).

Eligibility criteria

The inclusion criteria were as follows: (1) papers published between the 1st of January 2017 and the

Table 1 Brief presentation of the ML and DL models reported in the papers of our survey

Category	Model	Short description
ML	Random Forest [18]	Random Forest is an ensemble learning approach for classification and regression that works by generating a large number of decision trees during training. For classification problems, the output of a random forest is the class that most of the trees choose.
	Logit-Boost [19]	LogitBoost is a classification-boosting technique that performs additive logistic regression. In contrast to well-known boosting algorithms (such as AdaBoost, which reduces the exponential loss), LogitBoost minimizes the logistic loss.
	XGBoost [20]	XGBoost is a distributed gradient boosting toolkit tuned for efficiency, flexibility, and portability. It incorporates ML methods inside the context of Gradient boosting. XGBoost offers a parallel tree boosting (also known as GBDT, GBM) that efficiently and precisely solves several data science challenges.
	ANN [21]	ANNs provide a rough approximation of how the human brain processes and evaluates information. They are composed of successive layers, referred to as the input, hidden, and output layers. The hidden layer is responsible for processing and transmitting the information from the input layer to the output layer.
DL	CNN [12]	CNNs are utilized in computer vision and pattern recognition. CNNs are neural networks with the following layers: input, convolution, pooling, and fully connected layers. Convolution layers scan input dimensions using filters that execute convolution operations. Pooling is a down-sampling process used after a convolution layer. Fully connected layers, in which each input is connected to all neurons in the following layer, are commonly found towards the end of CNN designs to improve class scores.
	3D-CNN [22]	3D CNNs are a generalization of 2D CNNs in 3D. They accept a 3D volume or a series of 2D frames as input. Then, kernels traverse three dimensions of data to generate three-dimensional activation maps. They learn effective representations of volumetric data overall.
	Graph CNN [23]	Graph CNN is a convolutional neural network that can directly operate on graphs and use the structural information they contain. It addresses the issue of categorizing graph nodes when labels are supplied for just a limited portion of nodes (semisupervised learning).

15th of March 2022 (date of literature search)); (2) SPECT MPI images used for the CVD classification in nuclear medicine; (3) AI-based algorithms, including both traditional ML and DL techniques investigated for cardiovascular diagnosis; (4) explainable artificial intelligence of SPECT for diagnosis of coronary artery disease.

The exclusion criteria were as follows: (1) articles not related to AI/ML; (2) articles published before 2017; (3) Websites and online material; (4) nonoriginal research articles, such as protocols, reviews, meta-analyses, etc.; (5) articles not written in English; (6) student theses, book chapters, editorials, and commentaries; (7) papers describing frameworks, platforms, software, libraries, etc.; (8) papers investigated PET or PET/CT for cardiovascular diagnosis.

Literature collection

The literature search was performed by supplying the search strings for each database, as shown in section 3.2. As a response, these search queries returned many publications. The search results from each database were assessed according to the predefined inclusion/exclusion criteria. The outcomes of 342 research articles were considered based on the selection criteria. As per the exclusion criteria, only qualified articles, book chapters and summary reports were chosen. Journal editorials, newsletters, and papers not in English were excluded. Thesis reports and brief reports were excluded too. According to the inclusion criteria, we abided by the following considerations: reference of the author, year of publication, whether it belongs to a journal or conference proceeding, the definition of the relevant CAD diagnosis focusing on the type of AI methods, such as DL, ML, CNNs and explainable AI.

Furthermore, we scrutinized the abstract and summary of the chosen articles to investigate whether the selected articles fully satisfy the inclusion criteria. All insignificant and unrelated articles were discarded at this stage. Similarly, all academic research papers that did not match the inclusion criteria of AI methods were discarded too. After reviewing the remaining 82 papers, the second stage of articles' extraction followed, in which papers describing AI frameworks, platforms, software, nuclear medicine libraries and papers investigating PET or PET/CT for cardiovascular diagnosis were excluded.

Finally, 23 studies satisfied the eligibility criteria and were chosen for in-depth analysis and study. The complete process is visualized in Fig. 1.

Results

The research study identified 23 related publications that qualify for reporting. Fig. 2 shows the distribution of publications during the investigated period.

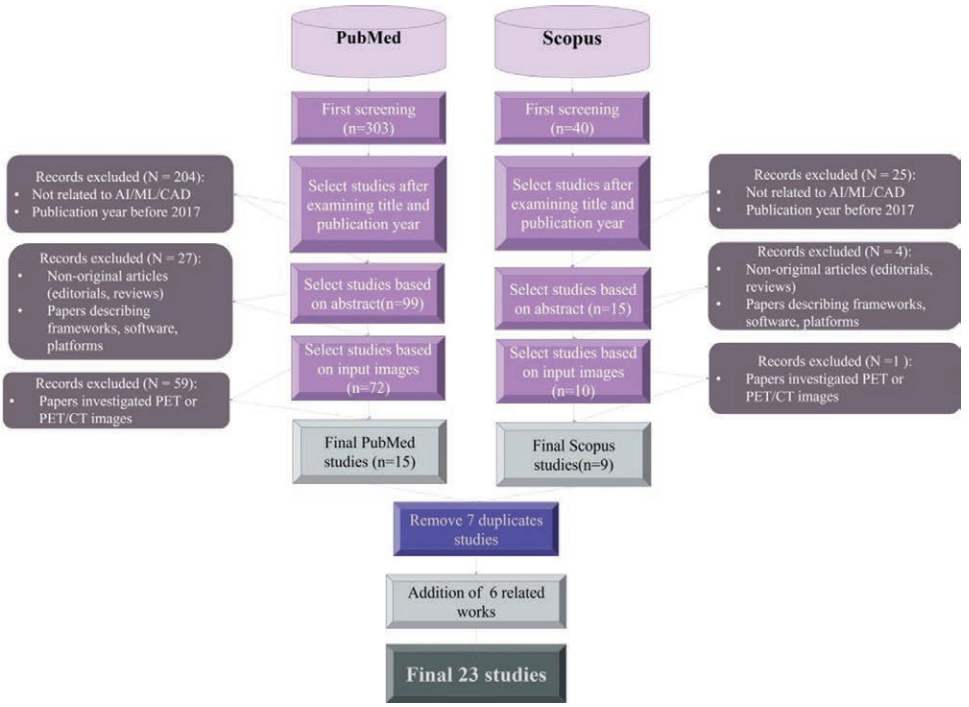
There is an evident increase in publications, reflecting the growing interest in AI-based CAD diagnosis in nuclear medicine utilizing SPECT MPI scans. As for the limited number of publications in 2022, this is since this review was conducted up to June 2022. However, an increasing tendency in the number of publications is anticipated. Out of the 23 reviewed publications, 22 are published in peer-reviewed journals. An analytical overview of the journals and conferences that participate in publishing is presented in Fig. 3.

We thoroughly reviewed all selected articles and finally retained those which applied AI methods such as ANN, SVM, Random Forest (RF), ML, DL, CNN, etc., for CAD diagnosis in nuclear medicine.

Related work in diagnosis/classification

Exploring the research conducted so far regarding computer-aided systems addressing CAD diagnosis with

Fig. 1



Literature search and qualification process.

SPECT MPI data, many published papers have applied CNNs and investigated their capabilities in classifying images. More specifically, Otaki *et al.* [1]. aimed to construct an explainable DL model to detect any signs of obstructive CAD in SPECT MPI images. For this research, 3.578 patients with suspected CAD were included in only stress representation. A model from scratch was developed, including age, sex, and cardiac volumes of patients in the last fully connected layer, so the model further understands CAD's characteristics. Concerning the results, the DL model provided 0.83% AUC after implementing 10-fold repeated testing to validate the accuracy, outperforming the reader's diagnosis, which produced 0.71% AUC. Moreover, attention maps were generated through the Grad-CAM (gradient-weighted class activation mapping) algorithm to highlight the parts that reveal the corresponding output.

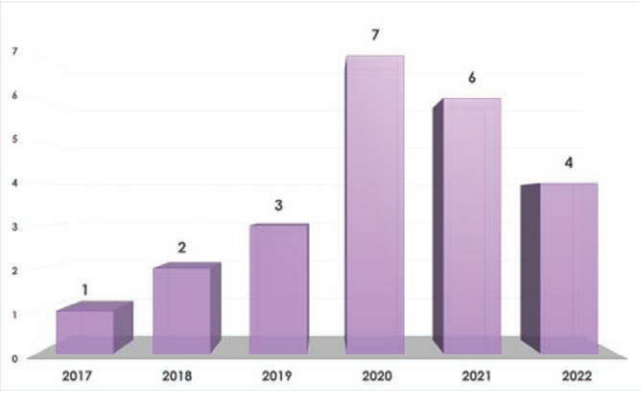
Chen *et al.* [8]. focused on implementing a three-dimensional CNN to classify MPI scans into abnormal and normal. A total of 979 patients were included in CZT (cadmium zinc telluride) SPECT format and greyscale mode. Also, Grad-CAM was applied to observe the parts contributing to the corresponding prediction. The 3D model extracted 87.64% accuracy using a five-fold cross-validation technique and shows great potential in assisting the clinical diagnosis of SPECT images.

Spier *et al.* [25]. developed a Graph CNN to automatically classify myocardial polar maps into normal and abnormal. The dataset includes 946 labeled images in a sitting upright position and stress and rest representation. Additionally, a four-fold cross-validation technique was applied to validate the extracted performance of the model. Moreover, heatmaps were produced, so the regions that indicate the abnormality and their frequency of appearance are demonstrated. The model achieved 89.9% on rest polar maps and 91.1% on stress polar maps, demonstrating that the model can support future clinical decisions.

Furthermore, Liu *et al.* [26]. applied a deep CNN, termed ResNet-34, to automatically diagnose myocardial perfusion abnormalities. A total of 37,243 patients were included in stress-only SPECT MPI representation and were labeled as normal or abnormal by a nuclear expert. In addition, six features were added as a second input, and these are gender, BMI, length, stress type, radiotracer and the option of including attenuation correction or not. The AUC value for DL (0.872) was higher than that of the standard quantitative perfusion defect size (DS) method (0.838). DL's performance was evaluated with a five-fold cross-validation approach.

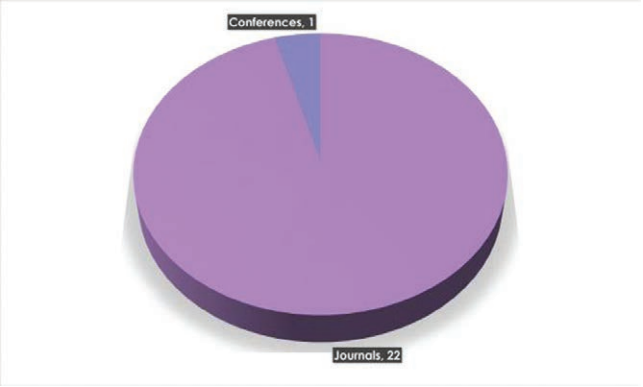
Berkaya *et al.* [5]. proposed two different classification models and a knowledge-based one for contributing to diagnosing perfusion abnormalities with SPECT MPI data. The first classification model uses transfer learning

Fig. 2



Number of publications per year.

Fig. 3



Number of publications in scientific peer-reviewed journals and conferences.

with various pretrained deep NNs. In contrast, the second is a robust classifier, SVM, which utilizes deep and shallow features extracted from the pretrained networks. Concerning the knowledge-based model, it aims at converting the knowledge of nuclear experts to image processing techniques. A total of 192 cases were used in stress and rest representation. The DL-based model extracted 94% accuracy, 88% sensitivity and 100% specificity, whereas the knowledge-based model achieved 93%, 100% and 86%, respectively.

Nakajima *et al.* [27]. proposed an ANN to evaluate the prediction of diagnosing CAD, in contrast to the conventional quantitation approach. The dataset consists of 1001 stress/rest images, and the possible outputs are abnormal and normal. Additional data were included, such as age, sex, height, weight, risk factors and results of coronary angiography and history of PCI (percutaneous coronary intervention) or CABG (coronary artery bypass grafting). Concerning the validation dataset, 364 cases were added to achieve generability and stability. Overall, ANN outperformed the conventional method, reaching 0.92% of the AUC value in contrast to the 0.82% that the conventional method extracted.

Otaki *et al.* [28]. developed a DL model for diagnosing obstructive CAD in polar maps. A total of 1160 patients were included, without known CAD, in upright and supine positions and only in stress representation. Also, clinical data were added, such as sex and BMI. Grad-CAM was utilized to visualize the regions responsible for disease prediction. The results showed greater sensitivity in the DL approach, achieving values of 82% in men and 71% in women. In comparison, the corresponding values for the standard SSS (Summed Stress Score) approach were 75% and 71%, for the upright (U-TPD) were 77% and 70% and for the supine (S-TPD) were 73% and 65% in men and women, respectively. Following the provided results, the study showed that the diagnostic performance of DL yielded notable differences from D-SPECT between men and women for the prediction of obstructive CAD. This might be derived from the fact that certain factors connected to sex, such as cardiac size, differ from men to women. In addition, leave-one-center-out external validation was performed to avoid overfitting.

A deep CNN to evaluate the automatic prediction of CAD compared to total perfusion deficit (TPD) was developed by Betancur *et al.* [29]. It has to be mentioned that sex information was added between the fully connected layers. A total of 1.638 patients without known CAD were included in the stress demonstration, where polar maps were acquired. DL achieved a higher AUC than TPD, which was 0.8% and 0.78% per patient case and 0.76% and 0.73% per vessel, respectively. Moreover, DL was evaluated using a stratified 10-fold cross-validation procedure to prevent overfitting. Next, Betancur *et al.* [30]. assessed the prediction of obstructive disease by applying CNN and comparing the produced results with those of the standard combined TPD method. A total of 1.160 polar maps were included in semiupright and supine stress representation. Moreover, sex information was added, so there can be a differentiation between men's and women's pattern extraction. The AUC values were 0.81% and 0.78% per patient and 0.77% and 0.73% per vessel for CNN and TPD, respectively, concluding the superior performance of CNN. For further validation, the novel leave-one-center-out procedure was applied.

Apostolopoulos *et al.* [31]. implemented a hybrid CNN-random forest model to diagnose CAD in polar map images, adding clinical data. Five hundred sixty-six patients were included in stress and rest demonstration, obtained from both attenuation-corrected (AC) and non-attenuation-corrected (NAC) cases. The clinical data include gender, age, symptoms, predisposing factors and recurrent diseases. Moreover, a data augmentation technique was applied due to the small dataset. Concerning

the classification task, InceptionV3 was developed, where the input image was classified as abnormal or normal. The output of this model was inserted into the FR and NN hybrid model, utilizing the clinical data to predict the final label. The proposed method achieved 79.15% accuracy, using a 10-fold cross-validation procedure to evaluate the model's performance thoroughly.

Additionally, Apostolopoulos *et al*. [32]. investigated deep CNNs for the automatic classification of polar maps for myocardial perfusion diagnosis. The corresponding dataset includes 216 stress and rest representation cases, including AC and NAC polar maps. To tackle the small size of the dataset, the authors applied two methodologies: transfer learning with VGG-16 and VGG-19 as pretrained networks, while they also utilized data augmentation. The produced results were compared with a standard semiquantitative polar map analysis. Pre-trained network VGG-16 achieved an accuracy of 74.53%, a sensitivity of 75%, and a specificity of 73.43%, whereas semiquantitative analysis extracted 66.20% and 64.81% accuracy for AC and NAC, respectively.

Recently, Papandrianos *et al*. [33]. investigated the abilities of CNN to diagnose ischemia or infarction in SPECT MPI scans. Two approaches were followed; the first includes the implementation of an RGB-CNN from scratch, and the second involves the application of transfer learning. This robust method includes pretrained networks, like VGG16, DenseNet, MobileNet and InceptionV3, which were applied to the corresponding dataset. The research comprises 224 patients who underwent a stress and rest procedure. Furthermore, the data augmentation technique was utilized to generate new images after applying range, enlargement, rotation, and flip techniques to increase the dataset's number. The overall accuracy is 93.47% ±2.81 with a 0.936 AUC value. Papandrianos and his team [14] also explored the application of CNN to SPECT MPI data for successfully classifying normal and abnormal cases regarding CAD. A total of 513 patients were included in stress and rest representation, with no additional clinical information. Also, the data augmentation technique was used to increase the small number of images in the training dataset, producing images through zoom and rotation. The CNN model extracted magnificent results with 0.937 AUC and 90.21% accuracy.

Zahiri *et al*. [34]. investigated CNNs and exploited their capabilities in classifying polar maps in SPECT MPI format for diagnosing CAD. The dataset includes 3.318 images in stress and rest representation, labeled as normal and abnormal, following an expert doctor's diagnosis. Furthermore, data augmentation was utilized to reduce overfitting and achieve generalization. The authors explored two approaches. The approach that refers to the stress and rest images acquired an AUC of 0.845, in contrast to the one that utilized only stress images, which achieved an AUC of 0.827. For the further evaluation of the produced model, a five-fold cross-validation procedure was followed.

Filho *et al*. [35]. employed RF, an ML algorithm, to classify normal and abnormal cases of patients. This study included 1.007 SPECT MPI stress/rest images, each divided into five vertical and horizontal slices generating ten attributes. Data augmentation was utilized. According to the extracted results, RF achieved an AUC of 0.853, an accuracy of 0.93, a precision of 0.968 and a sensitivity of 0.963. A 10-fold cross-validation technique was utilized to evaluate the results.

Trung *et al*. [36]. proposed VGG-16 as a CNN model for diagnosing CAD, using both SPECT and polar images. A total of 1413 heart SPECT images were included and were labeled by a nuclear expert as CAD or not. The VGG-16 performance was evaluated with five-fold cross-validation. According to the results, SPECT images provided higher accuracy, with a precision of 86.14 ± 2.14% and 82.57% ± 2.33% for SPECT and polar maps.

Arvidsson *et al*. [37]. developed a CNN model for the prediction of CAD. A total of 588 patients were available for this study, along with angina symptoms, age, BMI, pretest probability, ESC, and AHA probability as clinical data for each case. Moreover, data augmentation was performed. CNN achieved an AUC value of 0.89 at a per-vessel level, and 0.95 per patient, and the results were evaluated by implementing a five-fold cross-validation approach.

A ML approach named LogitBoost, which constitutes an ensemble boosting algorithm to predict MACE by combining clinical records and SPECT images was explored by Betancur *et al*. [38]. A total of 2619 patients were included in stress demonstration, along with clinical data like age, sex, and risk factors which were hypertension, diabetes mellitus, dyslipidemia, smoking and family history of CAD. The results were evaluated with a stratified 10-fold cross-validation methodology and compared against traditional methods like ML with all available data (ML-combined), ML with only SPECT images as input, expert diagnosis, and automatic quantitative analysis by stress and ischemic TPD. Overall, MACE prediction with ML-combined outperformed AUC 0.81, where expert diagnosis extracted AUC 0.65 and ML with images acquired AUC 0.78.

In the study by Rahmani *et al*. [39]., the authors demonstrated an ANN approach to predict CAD successfully. A total of 93 patients were included in stress and rest examination. The developed ANN made use of clinical and quantification data, which are reports of MPI, counts of 40 segments of stress/rest polar maps, age, gender and risk factors. The results indicated that ANN outperformed with 85.7% accuracy, adding age, gender, and cardiac risk factors as input variables.

Hu *et al.* [40]. inspected per-vessel and per-patient predictions by developing LogitBoost. The number of patients involved in this study is 1980 and went under stress and rest examination; in total, 18 clinical, 9 stress tests and 28 imaging variables were included. According to the results, LogitBoost outperformed nuclear cardiologists with per-vessel and per-patient AUC of 0.79 and 0.81, respectively, in contrast to stress TPD with an AUC of 0.71 and ischaemic TPD with an AUC of 0.72. Moreover, the ML approach was evaluated with 10-fold cross-validation to estimate model performance accurately.

Miller *et al.* [41]. developed an explainable DL-based model to improve the accuracy of MPI. A total of 240 patients in this study underwent an MPI examination. The DL approach for CAD diagnosis achieved an AUC of 0.779, compared to the expert's analysis with an AUC of 0.747 and stress TPD of 0.718. Also, the proposed DL method for CAD diagnosis extracted the best sensitivity overall. The authors developed an explainable methodology to provide meaningful interpretability.

Recently, Papandrianos and his team [42] developed an RGB-CNN and utilized SPECT images to classify patients between infarction, ischemia, and normal classes. A total of 647 patients were included in this study, and data augmentation was employed. A comparative study was conducted between the proposed model against pretrained networks VGG-16 and DenseNet-121. The proposed RGB-CNN model outperformed with an accuracy of 93.33%, while VGG-16 and DenseNet-121 achieved 88.54% and 86.11%, respectively. The proposed model was further evaluated by utilizing a 10-fold cross-validation approach.

Furthermore, Papandrianos *et al.* [43]. demonstrated developing an RGB-CNN model for automatically classifying polar maps. Three hundred fourteen polar maps were included in stress and rest representation and AC and NAC format. A comparative study was conducted between RGB-CNN and the robust pretrained model VGG-16. RGB reached 92.07%, and VGG-16 achieved 95.83% accuracy, respectively.

Baskaran *et al.* [44]. explored the inclusion of SPECT MPI data and clinical data to detect and perform the revascularisation of CAD by developing a boosting algorithm, XGBoost. A total of 716 patients were included in this study. The proposed boosting method achieved an AUC of 0.779, a sensitivity of 89.2%, and a specificity of 92.9% when five-fold cross-validation was performed to estimate the results. According to the results, BMI, age and angina severity are the most valuable clinical information to be included in the process of prediction and revascularisation of CAD. Table 2 gathers the reviewed research publications as reported in the literature.

Many AI-based approaches have already been successfully incorporated within the framework for CAD diagnosis in nuclear medical imaging. However, DL-based methodologies have taken tremendous steps toward performance in nuclear medical image analysis while they enhance state-of-the-art CAD diagnosis using SPECT myocardial perfusion imaging scans. Thus, they seem preferable and prescribe future trends for further development.

Discussion

Major findings

This review addresses the CAD problem by utilizing SPECT MPI data and applying several methodologies, primarily based on DL and CNNs. The results produced from the application of DL methods are superior in contrast to other conventional methods. Overall, CNNs outperformed quantitative methods as they acquired high values for metrics like AUC, accuracy, sensitivity and specificity (see Table 2). More specifically, the AUC, accuracy, sensitivity and specificity values were higher in DL methods than in other quantitative methodologies, such as TPD. In all the reported studies, DL methods particularly improved diagnostic accuracy by approximately 10%. For example, in [25], the AUC score for disease prediction by DL was higher than for TPD (per patient: 0.80 vs. 0.78; per vessel: 0.76 vs. 0.73). With the DL threshold set to the same specificity as TPD, both per-patient sensitivity and per-vessel sensitivity improved in the same manner. However, the contribution of additional data needs to be examined to explain the functionality behind the prediction of NNs and increase accuracy. Age, sex and BMI are some clinical data already being utilized since they provide more information about a patient's status.

Furthermore, DL incorporates transfer learning, which is a robust method. So far, various pretrained networks like VGG-16, ResNet-32 and DenseNet-121 have been trained in image classification, so these models are considered capable of providing generalization and being applied to a variety of complex problems.

Through the current systematic review, valuable outcomes have been extracted, as illustrated in the following graphs, and further analyzed, evaluated and discussed so that certain insights are elicited, highlighting the contribution of AI methods for CAD diagnosis in nuclear medicine.

The literature analysis regarding the utilization of various AI technologies dictates that the DL methods (CNN, VGG, ResNet, hand-crafted CNN) are the most popular choice (see Fig. 4), having almost 75% preference over the family of ML-based models. In particular, the DL methods and their combinations are the most used AI methods in classification tasks concerning CAD diagnosis (74%). This indicates that the two major categories, CNN and hand-crafted CNN, demonstrate dominance in almost all cases (65%) regarding AI-based nuclear cardiology using SPECT MPI scans.

Table 2 Overview of the reported literature (23 studies)

No.	Ref.	Year	#patients	Input	Learning algorithm	Output	Validation	Explainability	Results	
									Accuracy	AUC-Sensitivity-Specificity
1	[1]	2021	3,578	SPECT stress images, sex, age, cardiac volumes	CNN (hand-crafted)	Probability of obstructive CAD	10FCV (3,578 subjects)	Attention maps	N/A	AUC: 0.83% sensitivity: 84.3% specificity: 61%
2	[5]	2020	192	SPECT images -stress/rest	CNN (VGG-16)	Normal or Abnormal	training (66%), validation (17%), and test (17%) (Total num images: 192)	N/A	94%	sensitivity 88% specificity 100%
3	[8]	2021	979	CZT SPECT images (grey)	3D-CNN (hand-crafted)	Prediction of CAD	5FCV	Grad-CAM	87.64%	sensitivity 81.582% specificity 92.16%
4	[25]	2019	946	Polar maps	Graph CNNs (hand-crafted)	Normal or abnormal	Hold-out for localization and 4FCV for the entire dataset	indirectly via segmenting the images into sub-images	N/A	Rest sensitivity - 85.1% Rest Spec 85.31% Stress sensitivity – 85.7% Stress specificity 95.9%
5	[26]	2021	37,243	SPECT stress images – clinical information	ResNet-34	Normal or abnormal	5FCV (Total num images: 37,243	N/A	82.7% ± 2.5	AUC: 0.872% sensitivity 74.4% ± 4.2 specificity 84.9% ± 3.7
6	[27]	2017	1,001	SPECT images stress/rest – clinical information	ANN	Normal or Abnormal	Hold out: 1,001 training, 364 test (the test set is multicenter) (Total num images: 1,001)	indirectly via segmenting the images into sub-images	N/A	AUC: 0.89 - 0.92 for stenosis ≥50%. AUC: 0.58–0.72 for stenosis ≥ 75%
7	[28]	2020	1,160	Polar maps - sex and BMI	CNN	Normal or abnormal	leave-one-center-out external validation (4 centers in total) (Total num images: 1,160)	Grad-CAM	N/A	sensitivity 82% in men and 71% in women
8	[29]	2018	1,638	SPECT stress images – sex info	CNN (hand-crafted)	Normal - Abnormal and per-vessel	10FCV	N/A	N/A	Per-patient sensitivity improved from 0.79 (TPD) to 0.82 (DL) Per-vessel sensitivity improved from 0.64 (TPD) to 0.69 (DL)
9	[30]	2020	1,160	SPECT stress - sex info	CNN (hand-crafted)	per-vessel risk for CAD	Leave-One-Center- Out CV	N/A	N/A	Per-patient: AUC of 0.81 (DL) vs AUC of 0.78 (cTPD) Per-vessel: AUC of 0.77 (DL) vs AUC of 0.729 AUC (cTPD
10	[31]	2021	566	Polar Maps stress/rest – clinical information	CNN (Inception V3) + Random Forest	Normal and Abnormal Polar Map and CAD diagnosis with extra clinical factors	10FCV	N/A	78.44%	Sensitivity 77.36%, Specificity 79.25%
11	[32]	2020	216	Polar maps	CNN(VGG-16)	Normal or Abnormal	10FCV (216 subjects)	N/A	74.53%	sensitivity 75% and specificity 73.43%
12	[33]	2021	224	SPECT stress/rest	RGB-CNN (hand-crafted)	Normal or Abnormal	Hold out 85%-15% (244 subjects)	N/A	93.47%+- 2.81	AUC 0.936%
13	[34]	2021	3,318	Polar maps	CNN (5 conv layers- 2 fully connected)	Normal or Abnormal	5FCV (Total num images: 3,318)	N/A	75.62%	AUC: 0.8450% sensitivity: 0.7856% specificity: 0.7434%
14	[35]	2021	1,007	SPECT stress/rest	Random Forest	Normal or Abnormal	10FCV (Total num mages: 1,007)	N/A	93.8%	AUC: 0.853% Sensitivity: 0.963%
15	[36]	2020	1,413	SPECT stress/rest and Polar maps	CNN (VGG-16)	Normal or Abnormal	5FCV (Total num images: 1,413)	N/A	N/A	Precision on SPECT 86.14% ± 2.14% and Polar map 82.57% ± 2.33%

(Continued)

Table 2 (Continued)

No.	Ref.	Year	#patients	Input	Learning algorithm	Output	Validation	Explainability	Results	
									Accuracy	AUC-Sensitivity-Specificity
16	[37]	2021	588	Polar maps – clinical information	CNN (5 conv layers – 2 fully connected layers)	probability of CAD in the left anterior artery left circumflex artery	N/A	N/A	N/A	AUC per-vessel: 0.89 AUC: per-patient: 0.95
17	[38]	2018	2,619	SPECT – clinical information	LogitBoost	MACE risk score	Stratified 10FCV (Total num images: 2,619)	N/A	N/A	AUC: 0.81
18	[39]	2019	93	Polar images stress/rest – clinical information	ANN	Normal or Abnormal	14 cases for validation (Total num mages: 93)	N/A	85.7%	N/A
19	[40]	2020	1,980	18 clinical variables, 9 stress-test variables, and 28 imaging variables	LogitBoost	Per-vessel and per-patient	10FCV (Total num images: 1,980)	N/A	N/A	Per-vessel AUC: 0.79 Per-patient AUC: 0.81
20	[41]	2022	240	SPECT	DL model	Normal and Abnormal	(Total num images: 240)	N/A	N/A	AUC 0.779
21	[42]	2022	647	SPECT stress/rest	CNN (handcrafted)	CAD probability	10-fold cross-validation	N/A	93.33%	AUC: 0.94
22	[43]	2022	314	Polar maps	RGB-CNN	Normal or abnormal	5FCV (Total num images: 314)	N/A	92%	N/A
23	[44]	2020	719	SPECT – clinical information	XGBoost	Normal or Abnormal SPECT + revascularization	5FCV (Total num images: 719)	N/A	N/A	AUC: 0.779

The presented classification studies can be addressed based on their desired outcomes. A few studies aim to predict CAD; consequently, the networks are trained with reference to the ICA findings. Such studies may involve e SPECT, Polar Map classification or segment-to-segment classification to identify local and per-vessel defects. SPECT MPI imaging, subsequent diagnostic examinations (such as the treadmill exercise test) and clinical information cannot guarantee a precise diagnosis regardless of the ML/DL model's robustness. For example, in [31], the authors compared the result of the SPECT scan with the surgical findings, coming across that SPECT yielded an accuracy of 75% in CAD identification. This is an inherent limitation of the current imaging modality. Therefore, applying ML and DL methods to the images is anticipated to exhibit suboptimal results. Indeed, most studies report an accuracy of $82 \pm 5\%$.

Most studies train their models using human interpretation as their ground truth. Though human visual inspection and reporting are subjective and may vary from study to study, measuring the agreement between the model and the expert is crucial. The reported studies exhibit excellent agreement ratings. For example, in [5], the authors reported a 92% agreement rating, while other studies report even better scores [43].

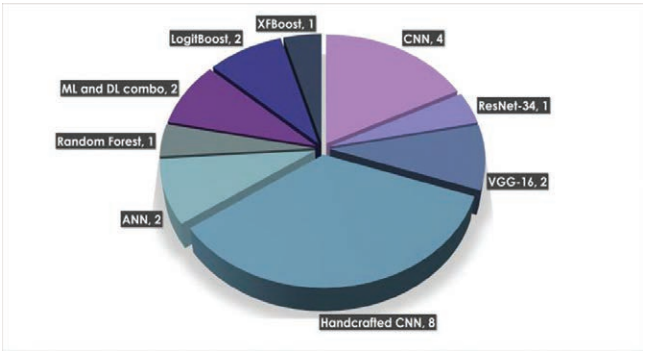
While summarising the key points and characteristics of the above-related literature, several limitations are extracted, such as the small number of datasets and the possibility of including bias that NNs may exhibit. Nevertheless, there has been adequate research regarding the potential that derives from the application of CNNs to the medical industry, which has offered magnificent results and close-to-expert analysis.

Most of the reported studies do not use methods to improve the explainability of their models. In general, NNs are labeled as 'black boxes' due to their unknown internal computational functionality; therefore, there is a hesitation to be fully adapted to nuclear imaging problems, and they are considered untrustworthy. This powerful limitation prevents them from being widely applied to medical problems until it is made clear that they are not biased. In the current literature, a few research studies were reported to implement XAI (explainable artificial intelligence) methodologies to observe the regions responsible for the corresponding prediction. In particular, two studies [7,14] employ the Grad-CAM [45] algorithm to identify essential regions in the image. In addition, providing separate model decisions on segments of the Polar Map, as presented in [25,27], improves the explainability of the model, though the model's reasoning is not entirely revealed.

Challenges

The domain of AI is considered unclear by humans regarding its behavior, especially in decision-making. Following

Fig. 4



ML and DL methods employed in the reported studies.

the literature review, XAI has been recently applied in nuclear imaging in a few research papers, particularly in CAD, to eliminate the model's bias. Therefore, there is a need for further research and experiment in this area. A recent challenge is to solve the interpretability and transparency issues in clinical nuclear medicine by investigating explainable AI methods based on deep learning and/or other ML techniques.

Integration of clinical factors can improve the diagnostic accuracy of the models and contribute to more explainable models. Even though some studies supply demographic characteristics to their models, such as the gender and sex of the subject, the importance of the demographic factors has not been examined further. In the study of Apostolopoulos *et al.* [31]., the authors used 23 clinical attributes and the DL model's prediction on the image to build an RF classifier. There was a 14% improvement in accuracy when the clinical attributes were embedded into the model. However, the importance of each attribute was not further examined.

Deep and robust DL methods require large-scale datasets for training. Otherwise, they tend to underfit, learn irrelevant features and become incapable of generalizing. Further, they are not robust to imaging device variations. Most reported studies utilize a small number of patient cases, usually deriving from a specific SPECT scanner. Multicentre and multidevice studies are mandatory in the future to assess the DL's efficiency in CAD diagnosis.

Furthermore, the medical domain requires secure transactions. On the one hand, data security would prevent the leakage of sensitive or confidential information. On the other hand, data integrity is also necessary to prevent corruption and information loss. Such issues may deal with using current technologies like blockchain, but certainly, this is a topic that deserves research.

Future studies are directed toward developing more robust techniques like data augmentation to resolve the issue of small datasets. Furthermore, more exploration is needed around CNN architectures, with the utilization of both clinical data and images and the contribution of explainable techniques, so the diagnosis provided by the CNN models follows the medical experts' diagnosis. Hence, they can provide a deeper understanding and more accurate results, overcoming the obstacle of black-box functionality.

Concluding remarks

This study reviews the current state of the art and existing results from research articles regarding DL-based classification algorithms in SPECT myocardial perfusion imaging for cardiovascular diagnosis. A systematic literature review protocol was chosen properly to meet the criteria of this research study. The research identified 23 studies. The most popular online and well-respected publishers and databases were considered to cover a range of disciplines, from the theoretical to the applied. A close look at the produced outcomes shows that CNN models have significantly contributed to heart disease diagnosis in nuclear medicine using SPECT MPI scans. Seeking further, explainable models are the current trend for automatic diagnosis and computer-aided systems to support nuclear medicine experts' decisions using AI methods.

Acknowledgements

The research project was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the '2nd Call for H.F.R.I. Research Projects to support Faculty Members & Researchers' (Project Number: 3656).

Conflicts of interest

There are no conflicts of interest.

References

1 Otaki Y, Singh A, Kavanagh P, Miller RJH, Parekh T, Tamarappoo BK, *et al.* Clinical deployment of explainable artificial intelligence of SPECT for diagnosis of coronary artery disease. *JACC Cardiovasc Imaging* 2022; **15**:1091–1102.
2 Strodthoff N, Strodthoff C. Detecting and interpreting myocardial infarction using fully convolutional neural networks. *Physiol Meas* 2019; **40**:015001.
3 Savvopoulos CA, Spyridonidis T, Papandrianos N, Vassilakos PJ, Alexopoulos D, Apostolopoulos DJ. CT-based attenuation correction in Tl-201 myocardial perfusion scintigraphy is less effective than non-corrected SPECT for risk stratification. *J Nucl Cardiol* 2014; **21**:519–531.
4 Mostafapour S, Gholamiankhah F, Maroufpour S, Momennezhad M, Asadinezhad M, Zakavi SR, *et al.* Deep learning-guided attenuation correction in the image domain for myocardial perfusion SPECT imaging. *J Comput Des Eng* 2022; **9**:434–447.
5 Kaplan Berkaya S, Ak Sivrikoz I, Gunal S. Classification models for SPECT myocardial perfusion imaging. *Comput Biol Med* 2020; **123**:103893.
6 Kawauchi K, Hirata K, Katoh C, Ichikawa S, Manabe O, Kobayashi K, *et al.* A convolutional neural network-based system to prevent patient misidentification in FDG-PET examinations. *Sci Rep* 2019; **9**:7192.
7 Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. *Artif Intell Rev* 2020; **53**:4093–4160.
8 Chen J-J, Su T-Y, Chen W-S, Chang Y-H, Lu HH-S. Convolutional neural network in the evaluation of myocardial ischemia from CZT SPECT myocardial perfusion imaging: comparison to automated quantification. *Appl Sci* 2021; **11**:514.

9 Ma L, Ma C, Liu Y, Wang X. Thyroid Diagnosis from SPECT Images Using Convolutional Neural Network with Optimization. *Comput Intell Neurosci* 2019; **2019**:1–11.
10 Hardesty L. Explained: neural networks. MIT News [Internet]. 14 April 2017. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414. [Accessed 15 October 2022]
11 Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995; **20**:273–297.
12 LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. *Handb Brain Theory Neural Netw* 1995; **3361**:1995.
13 Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, *et al*. Generative adversarial nets. *Adv Neural Inf Process Syst* 2014; **2**:2672–2680.
14 Papandrianos N, Feleki A, Papageorgiou E. Exploring classification of SPECT MPI images applying convolutional neural networks. In: 25th Pan-Hell Conf Inform [Internet]; Volos Greece: ACM; 2021. pp. 483–489. https://dl.acm.org/doi/10.1145/3503823.3503911. [Accessed 28 July 2022].
15 Goodfellow I, Bengio Y, Courville A. *Deep learning*. MIT press; 2016.
16 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**:436–444.
17 Ntakolia C, Diamantis DE, Papandrianos N, Moustakidis S, Papageorgiou EI. A lightweight convolutional neural network architecture applied for bone metastasis classification in nuclear medicine: a case study on prostate cancer patients. *Healthcare* 2020; **8**:493.
18 Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 1998; **20**:832–844.
19 Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann Stat* [Internet] 2000; **28**:337–407. https://projecteuclid.org/journals/annals-of-statistics/volume-28/issue-2/Additive-logistic-regression--a-statistical-view-of-boosting-With/10.1214/aos/1016218223.full. [Accessed 2022 August 1].
20 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min [Internet]; San Francisco, California, USA: ACM; 2016. pp. 785–794. https://dl.acm.org/doi/10.1145/2939672.2939785. [Accessed 1 August 2022].
21 Hassoun MH, others. *Fundamentals of artificial neural networks*. MIT press; 1995.
22 Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D deep learning on medical images: a review. *Sensors* 2020; **20**:5097.
23 Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Netw* 2019; **6**:11.
24 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021; **n71**:1–9.
25 Spier N, Nekolla S, Rupprecht C, Mustafa M, Navab N, Baust M. Classification of polar maps from cardiac perfusion imaging with graph-convolutional neural networks. *Sci Rep* 2019; **9**:7569.
26 Liu H, Wu J, Miller EJ, Liu C, Yaqiang L, Liu YH. Diagnostic accuracy of stress-only myocardial perfusion SPECT improved by deep learning. *Eur J Nucl Med Mol Imaging* 2021; **48**:2793–2800.
27 Nakajima K, Kudo T, Nakata T, Kiso K, Kasai T, Taniguchi Y, *et al*. Diagnostic accuracy of an artificial neural network compared with statistical quantitation of myocardial perfusion images: a Japanese multicenter study. *Eur J Nucl Med Mol Imaging* 2017; **44**:2280–2289.
28 Otaki Y, Tamarappoo B, Singh A, Sharir T, Hu L-H, Gransar H, *et al*. Diagnostic accuracy of deep learning for myocardial perfusion imaging in men and women with a high-efficiency parallel-hole-collimated cadmium-zinc-telluride camera: multicenter study. *J Nucl Med Soc Nucl Med* 2020; **61**:92–92.
29 Betancur J, Commandeur F, Motlagh M, Sharir T, Einstein AJ, Bokhari S, *et al*. Deep learning for prediction of obstructive disease from fast myocardial perfusion SPECT: a multicenter study. *JACC Cardiovasc Imaging* 2018; **11**:1654–1663.
30 Betancur J, Hu L-H, Commandeur F, Sharir T, Einstein AJ, Fish MB, *et al*. Deep learning analysis of upright-supine high-efficiency SPECT myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. *J Nucl Med* 2019; **60**:664–670.
31 Apostolopoulos ID, Apostolopoulos DI, Spyridonidis TI, Papathanasiou ND, Panayiotakis GS. Multi-input deep learning approach for cardiovascular disease diagnosis using myocardial perfusion Imaging and clinical data. *Phys Med* 2021; **84**:168–177.
32 Papathanasiou ND, Spyridonidis T, Apostolopoulos DJ. Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. *Hell J Nucl Med* 2020; **23**:125–132.
33 Papandrianos N, Papageorgiou E. Automatic diagnosis of coronary artery disease in SPECT myocardial perfusion imaging employing deep learning. *Appl Sci* 2021; **11**:6362.
34 Zahiri N, Asgari R, Razavi-Ratki S-K, parach A-A. Deep learning analysis of polar maps from SPECT myocardial perfusion imaging for prediction of coronary artery disease [Internet]. In Review; 2021. https://www.researchsquare.com/article/rs-1153347/v1. [Accessed 15 October 2022].
35 de Souza Filho EM, Fernandes F de A, Wiefels C, de Carvalho LND, dos Santos TF, dos Santos AASMD, *et al*. Machine learning algorithms to distinguish myocardial perfusion SPECT polar maps. *Front Cardiovasc Med* 2021; **8**:741667.
36 Trung NT, Ha NT, Thuan ND, Minh DH. A deeplearning method for diagnosing coronary artery disease using SPECT images of heart. *J Sci Technol* 2020; **144**:022–027.
37 Arvidsson I, Overgaard NC, Astrom K, Heyden A, Figueroa MO, Rose JF, *et al*. Prediction of obstructive coronary artery disease from myocardial perfusion scintigraphy using deep neural networks. In: 2020 25th Int Conf Pattern Recognit ICPR [Internet]; Milan, Italy: IEEE; 2021. pp. 4442–4449. https://ieeexplore.ieee.org/document/9412674/. [Accessed 28 July 2022].
38 Betancur J, Otaki Y, Motwani M, Fish MB, Lemley M, Dey D, *et al*. Prognostic value of combined clinical and myocardial perfusion imaging data using machine learning. *JACC Cardiovasc Imaging* 2018; **11**:1000–1009.
39 Rahmani R, Niazi P, Naseri M, Neishabouri M, Farzanefar S, Eftekhari M, *et al*. Precisión diagnóstica mejorada para la imagen de perfusión miocárdica usando redes neuronales artificiales en diferentes variables de entrada incluyendo datos clínicos y de cuantificación. *Rev Esp Med Nucl E Imagen Mol* 2019; **38**:275–279.
40 Hu L-H, Betancur J, Sharir T, Einstein AJ, Bokhari S, Fish MB, *et al*. Machine learning predicts per-vessel early coronary revascularization after fast myocardial perfusion SPECT: results from multicentre REFINE SPECT registry. *Eur Heart J - Cardiovasc Imaging* 2020; **21**:549–559.
41 Miller RJH, Kuronuma K, Singh A, Otaki Y, Hayes S, Chareonthaitawee P, *et al*. Explainable deep learning improves physician interpretation of myocardial perfusion imaging. *J Nucl Med* 2022; **63**.
42 Papandrianos NI, Feleki A, Papageorgiou EI, Martini C. Deep learning-based automated diagnosis for coronary artery disease using SPECT-MPI images. *J Clin Med* 2022; **11**:3918.
43 Papandrianos NI, Apostolopoulos ID, Feleki A, Apostolopoulos DJ, Papageorgiou EI. Deep learning exploration for SPECT MPI polar map images classification in coronary artery disease. *Ann Nucl Med* [Internet] 2022; **36**:823–833. https://link.springer.com/10.1007/s12149-022-01762-4. [Accessed 4 July 2022].
44 Baskaran L, Ying X, Xu Z, Al'Aref SJ, Lee BC, Lee S-E, *et al*. Machine learning insight into the role of imaging and clinical variables for the prediction of obstructive coronary artery disease and revascularization: an exploratory analysis of the CONSERVE study. Zirlik A, editor. *PLoS One* 2020; **15**:e0233791.
45 Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proc IEEE Int Conf Comput Vis* 2017; **1**:618–626.