Annals of Nuclear Medicine
https://doi.org/10.1007/s12149-022-01762-4

ORIGINAL ARTICLE

Deep learning exploration for SPECT MPI polar map images classification in coronary artery disease

Nikolaos I. Papandrianos[1] · Ioannis D. Apostolopoulos[2] · Anna Feleki[1] · Dimitris J. Apostolopoulos[3] · Elpiniki I. Papageorgiou[1]

Received: 12 March 2022 / Accepted: 9 June 2022

Abstract

Objective The exploration and the implementation of a deep learning method using a state-of-the-art convolutional neural network for the classification of polar maps represent myocardial perfusion for the detection of coronary artery disease.

Subjects and methods In the proposed research, the dataset includes stress and rest polar maps in attenuation-corrected (AC) and non-corrected (NAC) format, counting specifically 144 normal and 170 pathological cases. Due to the small number of the dataset, the following methods were implemented: First, transfer learning was conducted using VGG16, which is applied broadly in medical industry. Furthermore, data augmentation was utilized, wherein the images are rotated and flipped for expanding the dataset. Secondly, we evaluated a custom convolutional neural network called RGB CNN, which utilizes fewer parameters and is more lightweight. In addition, we utilized the k-fold validation for evaluating variability and overall performance of the examined model.

Results Our RGB CNN model achieved an agreement rating of 92.07% with a loss of 0.2519. The transfer learning technique (VGG16) attained 95.83% accuracy.

Conclusions The proposed model could be an effective tool for medical classification problems, in the case of polar map data acquired from myocardial perfusion images.

Keywords Deep learning · Cardiovascular diagnosis · SPECT MPI scans · Convolutional neural networks · Image classification

Introduction

Cardiovascular disease, also known as coronary artery disease (CAD), is one of the most serious health issues in medical history. CAD disease includes the blockage of the arteries resulting in the decrease of blood flow in the heart [1, 2]. Single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) is considered a great asset to nuclear medicine doctors. It offers heart representation in three-dimensional format and extracts spectacular accuracy in detecting the possibility of CAD in an efficient economic factor. After the extraction of images through the application of MPI, the quantitative parameters are demonstrated as a polar map. This method reduces the cost and time of unnecessary further examinations [1]. For the purpose of achieving the highest accuracy, patient's heart is imaged twice, as a stress and rest demonstration. Recently, it has been proven that quantitative data are a valuable aid into the decision and the evaluation of the identification of

✉ Nikolaos I. Papandrianos
npapandrianos@uth.gr

Ioannis D. Apostolopoulos
ece7216@upnet.gr

Anna Feleki
annafele1@uth.gr

Dimitris J. Apostolopoulos
dimap@med.upatras.gr

Elpiniki I. Papageorgiou
elpinikipapageorgiou@uth.gr

[1] Department of Energy Systems, University of Thessaly, Gaiopolis Campus, 41500 Larisa, Greece

[2] Department of Medical Physics, School of Medicine, University of Patras, Patras, Greece

[3] Department of Nuclear Medicine, University Hospital of Patras, 26504 Rion, Greece

Published online: 30 June 2022

CAD disease and can be considered as a second opinion to reinforce the resolution of complex circumstances [1–3].

As regards the process of SPECT MPI analysis, the algorithm evaluates the disease by the distinction between stress and rest images in two class classification problems. The classes are classified by the experts as normal or abnormal. The representation of the blood flow and its function in general appears in high contrast, and therefore, CAD systems are increasingly being utilized in scientific diagnostics and they have an irreplaceable role in the diagnosis of MPI images [3]. An automated algorithm for classifying CAD images is highly necessary for nuclear physicians, since the increasing number of cases put doctors under big pressure [4].

Deep learning (DL) alludes to a variety of methods belonging to the broader domain of machine learning (ML) and artificial intelligence (AI). DL has adjusted its technologies to medical industry, so they can learn patterns and representations of large data, especially imaging instances. More specifically, convolutional neural networks (CNNs) have been implemented in medical imaging due to the breakthrough applications they have developed to deal with image classification problems [5]. A standard artificial neural network consists of more layers of processing and classification of the data, in contrast to traditional approaches and typical machine learning algorithms. Due to this fact, artificial neural networks are a tremendous appliance in the solution of large and complex datasets [6].

In regard to CAD diagnosis in nuclear image analysis, many previous research papers have investigated the contribution of machine learning methodology into automatic classification [1, 6–11]. Recent CAD systems extract data in highly respected analysis similar to SPECT MPI, which is one of the most stable methods in cardiovascular data analysis. Combined with machine learning techniques, CAD systems are proposed as a robust tool to nuclear physicians to classify SPECT images automatically, without the addition of extra data. In the exploration of relative studies, DL methodologies and more specifically CNNs have been applied mostly for CAD diagnosis, with very promising results. The results obtained from previously conducted studies showed that CNNs could achieve accuracy close to that of expert clinicians, when compared to other approaches [8].

In the proposed study, the authors explored an algorithm to provide an automatic classification of myocardial perfusion imaging (MPI) data for the diagnosis of CAD. The algorithm is based on DL techniques. It involves the implementation of CNNs and the utilization of transfer learning, which have extracted remarkable results in medical imaging.

The aim of this research focuses on information extraction from SPECT data, classified by an expert reader. The innovation involves the contribution of a predefined RGB-CNN architecture, which has been concluded after a thorough exploration of parameters, and the proposal of a pretrained model, and more specifically VGG16, for improving CAD detection. The evaluation of the model is accomplished through reliable metrics and the usage of k-fold cross validation. Furthermore, we compare the proposed model to others in previous published studies [1].

Literature review on CNNs in nuclear medical image analysis

According to previous research papers, many efficient architectures have been proposed to classify polar maps for the detection of CAD. More specifically, in [9], Spier et al. concluded that graph CNNs can assist with the automatic classification of polar maps obtained from MPI, into normal and abnormal outputs. The decision of selecting graph CNNs was made because the polar maps are arranged in polar grid and graph CNNs do not need resampling of the data. Also, a fourfold cross validation was utilized for validating the performance of the proposed model, whereas data augmentation was not used. However, there are limitations due to the small training dataset and the fact that the expert reader that classified the data into two categories did not have access to the rest of the clinical information, and the procedure is not followed in a regular scenario. Nevertheless, the authors suggest that the proposed model could be an interesting option to help with clinical decisions by contributing with an automated algorithm.

Next, in [3], Betancur et al. studied the automatic prediction of obstructive disease comparing DL and total perfusion deficit (TPD) methodologies. The data for this research comprise raw and quantitative polar maps and a tenfold cross validation was utilized, so there can be a maximized usage of data. As a result, DL, and more specifically, the implementation of a CNN, demonstrated promising potentials to clinical methods, since the utilization of raw and quantitative polar maps seems to help this method to outperform TPD methodology. In addition, DL could achieve higher results if more data could be added.

Berkaya et al. [1] aimed at developing a computer-aided model to diagnose perfusion abnormalities. Two types of models were proposed. The first type uses transfer learning and the utilization of a support vector machine with deep features. The second type is based on the knowledge of expert clinicians. However, the limitations in this work are the same as the previous one: lack of data and lower accuracy if clinical data were not included. Regardless of the limitations, both types of this model can be utilized as an autonomous solution in computer-aided diagnosis systems.

Betancur in [7] evaluated the prediction of obstructive disease from upright-supine SPECT polar data in contrast with the standard method of total perfusion deficit (TPD). In the study, a fourfold leave-one-center-out cross-validation was utilized. After the extraction of results, it was observed that deep learning with MPI images outperformed TPD in the classification of obstructive disease. The deep learning approach was able to solve complex problems compared to other methods. The authors suggest that their method can easily adjust and deploy in the clinical industry as it is autonomous and only uses image data.

Apostolopoulos et al. [11] investigated the functionality of CNNs for automatic prediction of polar maps exported from MPI data for the diagnosis of CAD. The data included both attenuation-corrected (AC) and non-corrected (NAC) images. Due to the lack of training dataset, two methods were followed. The first method is transfer learning with the utilization of VGG-16 and the second is data augmentation, which includes rotation of images. Furthermore, a fivefold cross-validation was used to evaluate the model. The DL, as implemented by the authors, performed better than the standard quantitative methods. A second study by the same research team [10] verified the effectiveness of DL. In that study, the authors used a CNN to process the image data and a random forest classifier to process the clinical attributes. The results suggested that DL methods could be used as a part of a hybrid model that analyses both clinical information and imaging data.

Plenty of respected research works have been found in the related literature, however, there is a lack in the innovation of SPECT data, analyzed by convolutional neural networks. They are remarkably known for their ability to assist nuclear medicine with their outstanding results and automatically classify images. This study recommends a state-of-the-art CNN to be an integral tool in the nuclear industry.

Material and methods

CAD dataset

Patient data recorded in the department of Nuclear Medicine of the University Hospital of Patras were examined. The study covers a period from 30/3/2018 to 28/02/2020. Over this period, 1942 consecutive patients underwent gated-SPECT MPI with ^{99m}Tc-tetrofosmin. A hybrid SPECT/CT gamma-camera systems (Infinia, Hawkey-4, GE Healthcare) was used for MPI imaging. Computed tomography-based AC in both stress and rest images was applied in all cases. Data were acquired over 180° (from −45° up to + 135°). Three hundred and fifty-one patients were subsequently subjected to invasive coronary angiography (ICA) within 60 days from MPI for further evaluation. Thirty seven (37) patients were excluded from the dataset due to inconclusive MPI results or missing ICA reports. The final dataset of the present study includes 314 samples. Two nuclear medicine experts (N. Papandrianos and D. Apostolopoulos) were asked to label each instance of the dataset according to their expertise and experience. The nuclear medicine experts count several years of experience (15 and 30 respectfully). Labeling had been done by the experts using solely the Polar Map images of each patient. In this way, the model could be directly compared with the human experts. Hence, this study uses the experts' diagnostic yield as the ground truth and aims to furnish a DL model capable of competing with the human eye and the human expertise.

The Ethical Committee approved the study of our Institution. The nature of the survey waives the requirement to obtain patients' informed consent. The clinical characteristics of the dataset are presented in Table 1. Note that no missing values are reported for this particular study.

Research methodology

Convolutional Neural Networks (CNNs): a concise introduction

One of the most robust methods for image analysis is CNNs, which is a class of a deep neural network. More specifically, CNN consists of convolutional, pooling and fully connected layers for the purpose of extracting information from image data. This functionality has achieved remarkable results in a variety of deep learning projects [4].

First, there is the convolutional layer, where the operation of convolution is applied to the input and passes its result to the next layer so that the feature map is created. The feature map is an abstracted version of the image and contains pixel values, and it is computed with the following process. A kernel parses the complete width and height of the image

Table 1 Dataset characteristics

Clinical characteristics	Frequency
No	314
Age (mean ± sd)	65.3 ± 08.5 years
Sex (male/female)	73.88%/26.22%
History of CAD	37.89%
Previous Myocardial Infraction	32.48%
Previous Stroke	11.46%
Hypertension	83.12%
Dyslipidemia	65.6%
Smoking	53.18%
Diabetes	61.46%
Peripheral Angiopathy	9.23%
Family History of CAD	21.97%

and keeps the high-level features, or the patterns, which are the main characteristics of the differentiation of the classes.

Next, pooling layers follow. Their contribution is highly crucial for the CNN as it helps with the down-sampling of the image and removes unnecessary pixel values that increase the processing time of the neural network. It discards the pixel values that are higher than the given threshold and retains the values that are below the threshold. This benefits the reduction of the spatial size.

The last layer of the CNN is the fully connected layer, which applies to the flatten operation and converts the data into a vector and passes it to output, where the final label is given according to probabilities, which is based on activation functions [5].

Methodological framework

The CNN methodology includes five steps (Fig. 1): (a) collect data from MPI, (b) prepare data by normalizing theirs values, (c) CNN training, (d) CNN validation, and (e) evaluation of the model with the extraction of results. The proposed five steps are depicted in Fig. 1.

Step 1: Collect data

The MPI data given by the nuclear physicians are demonstrating the quantitative data of the heart's state in stress and rest, also known as polar maps. The expert physicians have given to each polar map a label and in the next step, each image was named accordingly to the given label, so as to be passed through the algorithm.

Step 2: Prepare data

Step 2.1: Data normalization and preparation.

Data normalization is a technique that is used in every machine learning algorithm, right before the training process. This method eliminates any outliers of the given values that are included in the dataset. For the polar map images, the pixel values, which initially varied from 0 to 255, where transformed into [0,1] space. Four (4) polar maps have been used in this study. Attenuation correction (AC) and non-attenuation correction (NAC) images in stress and in rest condition are concatenated into a single image, as illustrated in Fig. 2. The final image is of 720 × 680 size (width, height) and contains 32 bits per pixel. The concatenation of the four images into a single new image has proven to be effective for CAD diagnosis in recent literature [10, 11]. In this way,

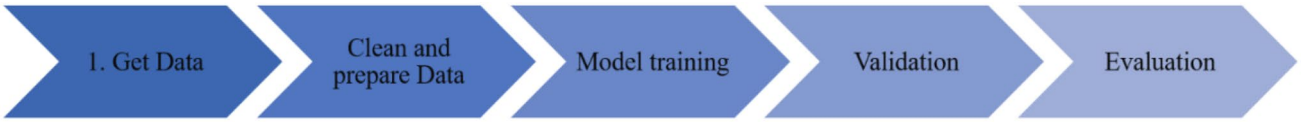

Fig. 1 Five steps of the recommended methodology

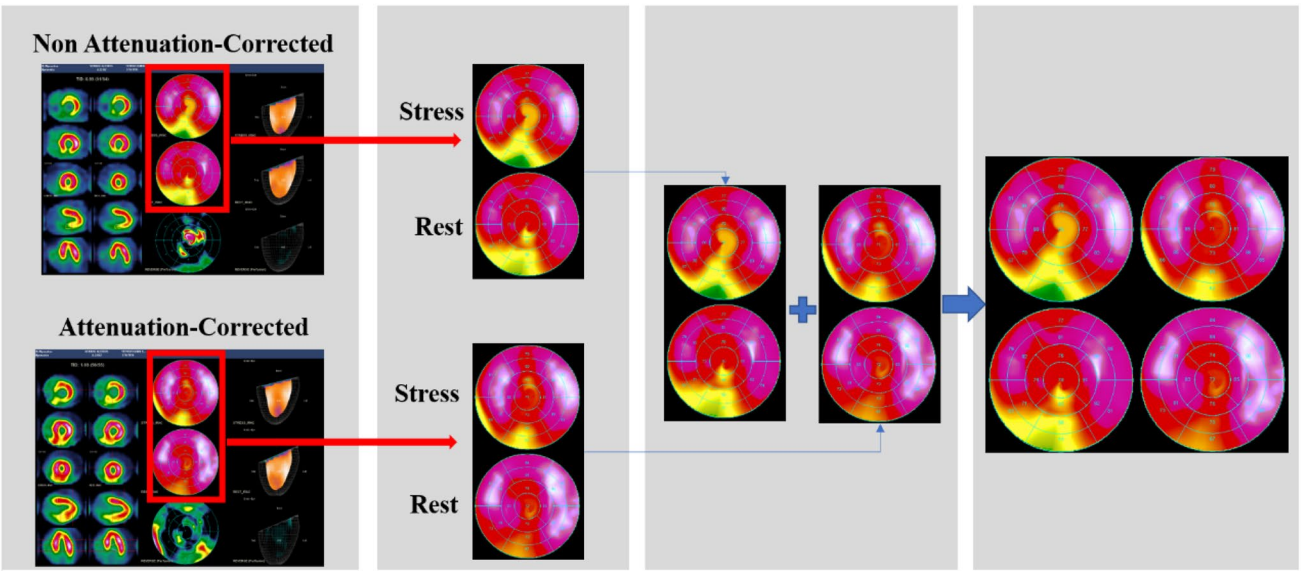

Fig. 2 Data preprocessing. Attenuation correction (AC) and non-attenuation correction (NAC) polar maps in stress and in rest condition are combined to construct the final image used by the proposed model

the model is given the opportunity to process all available information.

Step 2.2: Data shuffle

Our data pass through the training process in an alphabetical order, since they were divided in two categories, as it is a two-classification problem, where the labels are normal or abnormal. It should be noted that the data shuffle method was chosen, since data need to be selected randomly for the training process, so as the data would be processed randomly.

Step 2.3 Data split

Dataset split is a mandatory technique for extracting the best result and is conducted in three parts: validation, training and testing. Each part is being split with a specific percentage of the corresponding dataset. More specifically, 15% is for testing and the rest 85% is for the rest two parts, meaning 80% for training and 20% for validation.

Step 3: Training

Step 3.1: Data augmentation

The data augmentation technique is really important, when a small dataset is involved, otherwise, the CNN model would not be effective to classify other types of problems and would lack generalization capabilities. Data augmentation includes rotation, zooming, flipping and rescale of images [12]. In this way, the model becomes more robust and we circumvent the issue of data imbalance.

Step 3.2: Define CNN architecture

For defining the proper architecture of a CNN model, a detailed exploration of values of the parameters is needed, like pixel size, batch size, drop rate and number of nodes and layers of convolutional and dense layers. The exploration includes a few runs for each combination of values to conclude to the best one.

Step 3.3: Define activation function of CNN

The CNN model includes two types of functions: activation and loss functions. As regards the activation function, it specifies the output of the layers whereas the loss function is responsible for calculating the error of the prediction of the model. In most cases, RELU [13] is being selected as the activation function for convolutional and dense layers. As for the output layers, sigmoid for two class classification problems and softmax [14] for multi-output problems are selected. With respect to the loss functions, binary cross entropy is considered in the case of a two-class classification problem while in multi-output classification problems, categorical cross entropy is accounted [15].

Step 3.4: Train CNN

Training of a CNN model implies the insertion of the data through the model with the corresponding output and the extraction of a prediction. Then, the comparison with the real output follows and in the end, the back-propagation technique is applied, so that the weights can be adjusted to the corresponding dataset and more specifically to the output [16].

Step 4: Validation

Validation is an extremely useful procedure, in which parameters are being tuned while the algorithm is being evaluated to known data. In this way, a high accuracy to our responding dataset is achieved and overfitting [16] is avoided too. Validation is performed during the training stage. More specifically, we select a proportion of the training data that is held out for validation, during each training routine. In this way we can inspect the incremental learning process during each epoch, by calculating the validation data loss. As a result, the final trained model is ready to process unseen data from our test sets, because it has learned better what to expect from an unseen set in the first place.

Step 5: Evaluation

The evaluation process is performed by utilizing the testing set, after data split, which is unknown to the trained model. Throughout testing, our classifier is assessed, in terms of its ability to classify correctly and calculate the error of the false labeling. In the related research work, a number of known metrics is being used for the evaluation method like sensitivity, specificity, AUC value, precision, recall and f1-score for supplementary testing [17].

The proposed CNNs

The goal of this study is to implement an RGB-CNN model as a powerful tool to identify CAD in SPECT images, as it has shown tremendous progress in image recognition, with high accuracy and minimized loss. In order to conclude to the final architecture, a thorough exploration was conducted with different values for all parameters, such as pixels, epochs, batch size and number of nodes and layers. The definite architecture consists of four convolutional layers, each one of them is followed by a max pooling layer and a dropout layer. Next, the flatten layer is present to vectorize our data. Two dense layers follow, and finally, the output layer determines the label for each instance.

The recommended architecture resizes its input to 200×200 pixel size in RGB mode with 32 as batch size and each network runs for 200 epochs. Then, four convolutional layers follow with max pooling and dropout layers. The first convolutional layer includes two nodes, the second layer four nodes, the third layer eight nodes and the fourth layer sixteen nodes and all of them have 3×3 kernel size and ReLU as activation function. Furthermore, the max-pooling size is 2×2 and the drop-out rate is 0.2. Dropout [18] is an essential method to avoid overfitting. Briefly, this method randomly disconnects some of the extracted features from the classifier to urge the network to focus on a combination of features rather than sticking with some of them. Besides the convolutional layers, there is the flatten layer which converts the

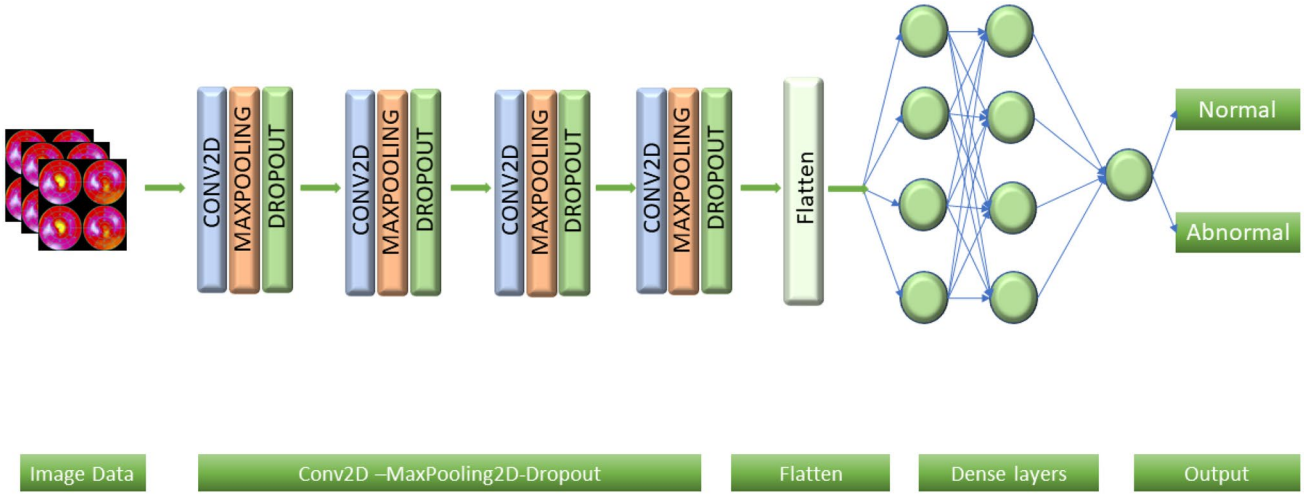

Fig. 3 The architecture of RGB CNN

data to a vector for down-sampling the image and boost the computation time with an addition of a dropout layer. Next, the fully connected layers with sixteen nodes each, follow and utilize ReLU as an activation function, so as to ensure non-linearity and capture the high-features of the image. Besides, ReLU activation function is the standard method for similar problems. The last component is the output which entails one node, as this is a two-class classification problem, employing the sigmoid activation function, as it seems in Fig. 3. We argue that a threshold set at 0.5 is optimal for the final decision. In essence, the sigmoid function outputs its predictions into the space (0,1). Values above the threshold are considered to correspond to the CAD class, whereas values below the threshold correspond to the Normal class.

The method of transfer learning is preferred instead of implementing a CNN from scratch, because the trained CNN has been trained previously on another use-case and is considered more stable to extract knowledge for another task. The pre-trained CNN that was utilized for this research work is VGG-16 and is generally used for nuclear medicine classification tasks for the reason that it achieves high results [10, 11, 19, 20]. VGG was able to extract valuable information from our dataset and then it was connected to our standard network, which was constructed based on the specific problem to complete the training, before the global average pooling method was added to down-sample the data. The only parameter we had to choose was the number of layers in VGG to achieve the best performance. We concluded that we would use 13 layers of VGG and then the extracted data would be added to our network with two fully connected layers with sixteen nodes each and using ReLU as the activation function. After every fully connected layer, a dropout layer followed to lower the spatial size. Finally, in the output, each instance of the dataset is being classified to a label, with Sigmoid as the activation function.

Due to the small number of images, model training was evaluated through a fivefold-cross-validation procedure. Cross-validation refers to the functionality in which the initial dataset is split into 5 chunks. The first 4 participate in the training stage and the fifth is used for the testing. This process is repeated until all chunks of the dataset are processed and used as the test set. A representation of a fivefold-cross-validation process can be seen in Fig. 4.

CNN demands hardware of high specifications; thus the experiment took place in a Cloud environment that Google facilitates and is named as Google Colab [21]. The CNN model and mainly transfer learning was implemented in Python 3.6 with libraries Keras 2.4.0 and Tensorflow 2.6.0. Furthermore, the operating system that was utilized, has the following specifications. (i) Processor: Intel® CoreTM i9-9980HK CPU 2.40 GHz, (ii) GPU: NVIDIA Quadro RTX 3000, (iii) Ram: 32 GB, (iv) System type: 64-bit operating system, × 64-based processor.

In order to evaluate our model, specific metrics were selected to demonstrate the abilities of the proposed architecture, in regard to the given dataset. These metrics are accuracy, loss, AUC value, ROC curve, sensitivity and specificity. More specifically, accuracy is the ratio between the correct labels that the algorithm classified and all images. Next, we have loss that demonstrates the error of the algorithm. The higher the loss the more deviation our dataset has to the proposed model. The AUC value is the representation of how well the model can correctly classify the images and the ROC Curve is the demonstration of this specific value. As regards sensitivity, it is the percentage of true positives whereas, specificity is the percentage of true negatives [22].

This research work investigates the analysis of a two-class classification problem. For this reason, the authors developed an RGB-CNN from scratch and also utilized transfer learning with VGG-16. Furthermore, data augmentation and fivefold cross validation were used for ensuring the efficiency of the proposed models.

Results

The following sector gathers the results that have been extracted after CNN classification. The results are the average values after 10 runs.

A thorough exploration was performed by testing different values for parameters that were included in a CNN architecture. In detail, different values for pixels were evaluated such as 200 × 200 × 3, 250 × 250 × 3 and 300 × 300 × 3, different values for batch sizes 8, 16, 32, 64 and also different

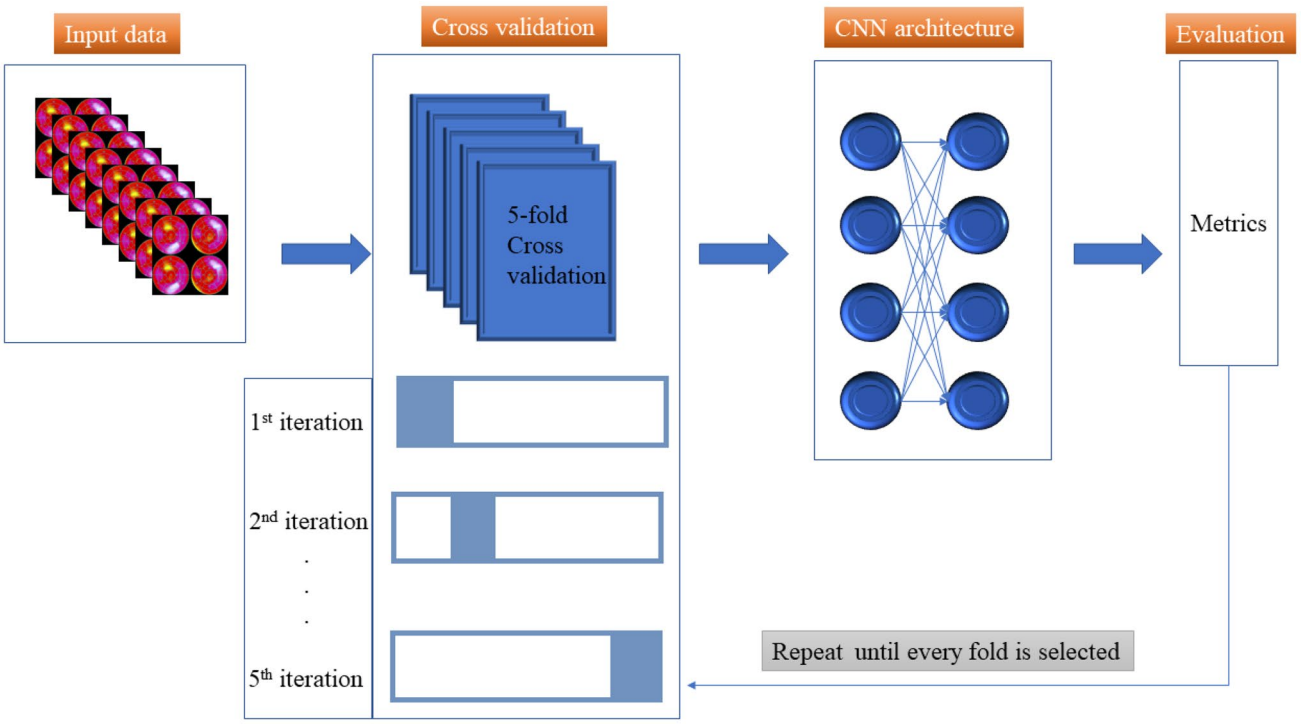

Fig. 4 Demonstration of fivefold cross validation

Table 2 Comparison of results between of different batch sizes

Batch size	Validation accuracy (%)	Validation loss	Test accuracy (%)	Test loss	AUC score
8	87.4046	0.2	91.2	0.25	0.91
16	87.406	0.3	90.4	0.23	0.90
32	94.443	0.2	92.7	0.22	0.92
64	87.033	0.3	88.5	0.23	0.89

Table 3 Comparison of results among different pixels

Pixels	Validation accuracy (%)	Validation loss	Test accuracy (%)	Test loss	AUC score
200 × 200 × 3	94.44	0.2	92.7	0.22	0.92
250 × 250 × 3	90.74	0.3	89.0	0.28	0.89
300 × 300 × 3	87.77	0.3	90.4	0.31	0.91

Table 4 Comparison of results between different number of convolutional layers and nodes

Layers	Validation accuracy (%)	Validation loss	Test accuracy (%)	Test loss	AUC score
2–4–8	91.35	0.2	87.5	0.30	0.88
2–4–8–16	94.44	0.2	92.7	0.22	0.92
2–4–8–16–32	89.26	0.3	90.8	0.22	0.91

Table 5 Comparison of results between of different number of dense nodes

Dense	Validation accuracy (%)	Validation loss	Test accuracy (%)	Test loss	AUC score
8–8	92.12	0.24	87.5	0.28	0.88
16–16	94.44	0.2	92.7	0.22	0.93
32–32	92.59	0.22	90.8	0.24	0.90
64–64	93.05	0.19	91.1	0.19	0.91

numbers of nodes and layers for convolutional layers, like 2–4–8, 2–4–8–16, 2–4–8–16–32 and for dense layers, like 8–8, 16–16, 32–32 and 64–64.

At the beginning, the selection for the parameters was random, until we concluded to 200 × 200 pixels, 2–4–8–16 for convolutional layers, and 16–16 for dense layers, whereas we differentiated batch sizes as depicted in Table 2. The following tables refer to performance metrics following a random train-test split, as explained in the previous sections.

After the exploration, 32 turned to be optimal batch size. The next step was to investigate which combination of image size is the best for the given dataset, as seen in Table 3.

The size of 200 × 200 × 3 (width, height, channels) seems to be the best image size, as it corresponds to the highest accuracy and the lowest loss. Next, we demonstrate the exploration process in determining the number of nodes and layers of convolutional layers that will extract the best results in Table 4.

The best selection is four layers and 2–4–8–16 nodes according to each layer. At this point, we have concluded that the best architecture is 200 × 200 × 3 pixel, 32 batch size and 2–4–8–16 for convolutional layer. The final step involves the exploration regarding the number of layers and nodes for dense layers in Table 5.

It is clear that 16–16 produces the best results, defining the best overall architecture. Accuracy, loss and ROC curve for the best architecture of RGB CNN are demonstrated in Fig. 5.

In Table 6, the average value of validation accuracy and loss, testing accuracy and loss, sensitivity and specificity are depicted accordingly.

The above-mentioned experiments guided us to define the optimal model parameters. Those parameters were used to perform a tenfold cross-validation. Next, in Table 7, we

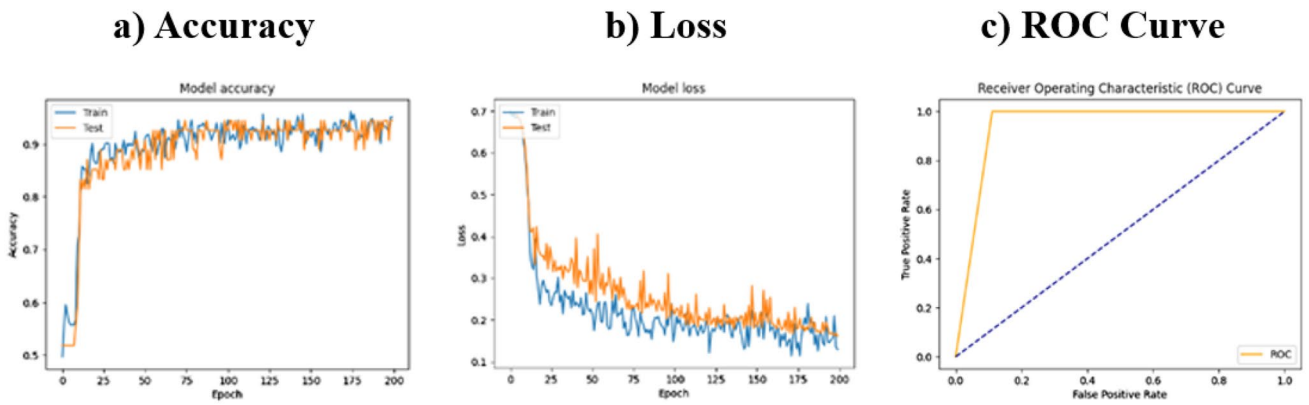

Fig. 5 Precision curves demonstrating the best RGB-CNN model in terms of accuracy, loss and ROC curve

Table 6 Sensitivity and specificity of best architecture of RGB-CNN

Validation accuracy (%)	Validation loss	Test accuracy (%)	Test loss	AUC score	Sensitivity	Specificity	Computation Time (s)
94.44	0.2	92.7	0.22	0.92	0.94	0.92	448

Table 7 Runs of tenfold cross validation for the proposed RGB-CNN

Folds	Accuracy	Loss
1	92.59	0.17
2	96.30	0.10
3	90.74	0.28
4	75.93	0.45
5	77.77	0.39
6	92.59	0.17
7	98.10	0.03
8	96.29	0.09
9	98.15	0.08
10	92.3	0.17
Average	91.08	0.19

present the results of the tenfold cross validation after data augmentation.

In the following (Table 8), we illustrate the concluding results of the best architecture for VGG-16 whereas, in Fig. 6, the plots that were generated for accuracy, loss and ROC curve are depicted. The best architecture for VGG-16 is 200 × 200 × 3 pixel size, 16 batch size, 13 trainable layers from VGG-16 default architecture and two fully connected layers with 16 nodes each.

For evaluating the outcomes and significance of this work, we proceeded with a comparison of the produced results with those of a similar research study that has been previously proposed, such as [7]. In that research study, the potential of VGG-16 being a robust factor in medical imaging was investigated, through the implementation of a fivefold cross validation and data augmentation. Taking one step further, we conducted the same experiment with our own CNN algorithm and dataset. It emerged that the proposed algorithm outperformed their model in terms of the produced results, as it is shown in Table 9. The computation time for the proposed VGG is 837 s in Colab. It is obvious that our algorithm provided outstanding results and can be considered as a more stable and robust method, as it is shown in Fig. 6.

Table 9 Comparison of results between two research studies

Models	Accuracy	Sensitivity	Specificity
Proposed VGG	95.83	0.97	0.94
Comparative VGG	74.53	0.75	0.73

Table 8 All runs from VGG best architecture

Runs	Validation accuracy	Validation loss	Test accuracy	Test loss	AUC score
1	97.91	0.10	97.91	0.18	0.98
2	95.83	0.45	95.83	0.49	0.96
3	97.92	0.11	95.83	0.13	0.96
4	91.66	0.19	93.75	0.20	0.94
5	95.83	0.05	95.83	0.07	0.95
Average	95.83	0.18	95.83	0.21	0.96

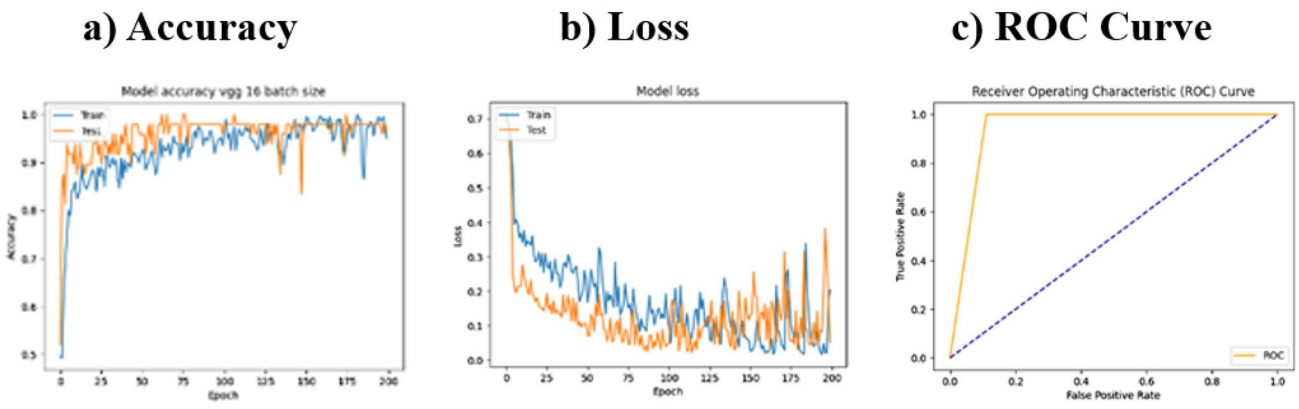

Fig. 6 Precision curves demonstrating the best VGG model in terms of accuracy, loss and ROC curve

Discussion

The experiments demonstrate the superiority of DL methods for interpreting MPI polar map images towards the diagnosis of Coronary Artery Disease. Both VGG-16 and RBG-CNN models are capable to accomplish this task, yielding remarkable results. The study used the experts' interpretation results as ground truth. Hence, we argue that both networks are showing great agreement rating with the experts (96%). It is shown that DL methods extracted significant image features to learn from the input data and provide a reasonable outcome.

This study has some limitations. First of all, only image data have been used. In the everyday practice, the nuclear medicine experts take into account both MPI results and other clinical data, such as the age, the sex, and predisposing factors. Hence, this study lacks completeness on this front. Future opportunities include the development of hybrid models that process both image and clinical information. Secondly, this study evaluated the DL effectiveness in interpreting the MPI results and not actually predicting CAD, as explained earlier. This is not to be considered as a limitation only, but also as a preliminary step towards a complete system.

Next, RGB-CNN, which is a custom-made CNN yields equivalent results with the state-of-the-art VGG-16 for this task. This is a remarkable feature and suggests that it is possible to employ lightweight and less complex CNNs to avoid computational costs and reduce the general complexity of the framework.

Finally, the parameter tuning experiments demonstrate that the combination of parameters affects the model's performance greatly. The optimal number of convolutional layers, dense layers, and the batch size improved the classification accuracy by at least seven (7) percent. This can also be considered as a limitation of custom-made DL models, because they require extensive parameter-tuning experiments. However, once their optimal parameters are defined and the training is complete, such models are ready for deployment in the everyday routine, without any other requirement.

Conclusions

Due to their previous successful results and their overall performance, CNNs were utilized in this two-class classification problem to classify CAD from SPECT data, into normal and abnormal. The dataset in the proposed study is small and no extra data were added. For this reason, the data augmentation technique was applied. Various architectures were tested in order to find the best algorithm for our corresponding dataset but with respect to generalization. The model was evaluated thoroughly with the fivefold cross-validation procedure and the produced results are promising. The proposed model could be a beneficial tool to automatic classify MPI images and can be a major asset for nuclear physicians.

Funding The research project was partially 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).

Data availability The dataset is not publicly available due to ethical reasons. However, the dataset is available to selected researchers on a reasonable request.

Declarations

Conflict of interest The author declares that they have no conflict of interest.

Submission declaration and verification The work described has not been published previously.

Data ethics The present study is retrospective.

References

1. Kaplan Berkaya S, Ak Sivrikoz I, Gunal S. Classification models for SPECT myocardial perfusion imaging. Comput Biol Med. 2020;123: 103893.
2. Cassar A, Holmes DR, Rihal CS, Gersh BJ. Chronic coronary artery disease: diagnosis and management. Mayo Clin Proc. 2009;84:1130–46.
3. Zhang YC, Kagen AC. Machine learning interface for medical image analysis. J Digit Imaging. 2017;30:615–21.
4. Arsanjani R, Xu Y, Dey D, Vahistha V, Shalev A, Nakanishi R, et al. Improved accuracy of myocardial perfusion SPECT for detection of coronary artery disease by machine learning in a large population. J Nucl Cardiol. 2013;20:553–62.
5. Stefanini M, Lancellotti R, Baraldi L, Calderara S. A Deep-learning-based approach to VM behavior Identification in Cloud Systems. In: Proceedings of 9th International Conference on Cloud Computing Service and Science [Internet]. Heraklion, Crete, Greece: SCITEPRESS—Science and Technology Publications; 2019 [cited 2022 Jan 15]. p. 308–15. https://doi.org/10.5220/0007708403080315
6. 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;60:664–70.
7. 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. JACC Cardiovasc Imaging. 2018;11:1654–63.
8. Papandrianos N, Papageorgiou E. Automatic diagnosis of coronary artery disease in SPECT myocardial perfusion imaging employing deep learning. Appl Sci. 2021;11:6362.

9. 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.
10. 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;84:168–77.
11. Apostolopoulos ID, 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;23:125–32.
12. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6:60.
13. Agarap AF (2019) Deep Learning using Rectified Linear Units (ReLU). ArXiv180308375 Cs Stat [Internet]. http://arxiv.org/abs/1803.08375. Accessed 15 Jan 2022
14. Hinton GE, Salakhutdinov RR. Replicated softmax: an undirected topic model. Adv Neural Inf Process Syst. 2009;22:1607–14.
15. Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for image restoration with neural networks. IEEE Trans Comput Imaging. 2017;3:47–57.
16. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436–44.
17. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. Mol Pharm. 2017;14:4462–75.
18. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1929–58.
19. Apostolopoulos ID, Papathanasiou ND, Panayiotakis GS. Classification of lung nodule malignancy in computed tomography imaging utilising generative adversarial networks and semi-supervised transfer learning. Biocybern Biomed Eng. 2021;41:1243–57.
20. Apostolopoulos ID, Pintelas EG, Livieris IE, Apostolopoulos DJ, Papathanasiou ND, Pintelas PE, et al. Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques. Med Biol Eng Comput. 2021;59:1299–310.
21. Bisong E. Google Colaboratory. Build mach learn deep learn models google cloud platf. Berkeley: Apress; 2019. p. 59–64. https://doi.org/10.1007/978-1-4842-4470-8_7.
22. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. Indian J Ophthalmol. 2008;56:45.