Proceedings of the Seventh International Conference on Research in Intelligent and Computing in Engineering pp. 261–266 ISSN 2300-5963 ACSIS, Vol. 33
DOI: 10.15439/2022R07

A Deep Learning Approach with Stack of Sub-classifiers for Multi-label Classification of Obstructive Disease from Myocardial Perfusion SPECT

Ninh Ngan Trieu
Faculty of Information Technology
Le Quy Don Technical University
Hanoi, Vietnam
trieuninhngan2012@gmail.com

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

Chi Thanh Nguyen*
Institute of Information Technology
AMST
Hanoi, Vietnam
thanhnc80@gmail.com
*Corresponding author

Thanh Trung Nguyen
Department of Medical Equipment
108 Military Central Hospital
Hanoi, Vietnam
thanhtrungys@yahoo.com

***Abstract*—Artificial intelligence applications, especially deep learning in medical imaging, have gained much attention in recent years. With the computer's aid, Coronary artery disease (CAD) - one of the most dangerous cardiovascular diseases - is diagnosed effectively without human interference and efforts. A lot of research involving predicting CAD from Myocardial Perfusion SPECT has been conducted and given impressive results. However, all existing methods detect whether there is a disease or not. They do not provide information about which obstructive areas are (mainly in the left anterior descending artery (LAD), left circumflex artery (LCx), and right coronary artery (RCA) territories) that result in CAD. To further diagnose CAD, we develop new classifiers to solve a multi-label classification problem with the highest accuracy and area under the receiver operating characteristics curve (AUC) when compared to different methods. Our proposed method is based on transfer learning to extract features from Myocardial Perfusion SPECT Polar Maps and a novel stack of sub-classifiers to detect particularly obstructive areas. We evaluated our methods with eight hundred and one obstructive images from a database of patients referred to a hospital from 2017 to 2019.**

***Index Terms*—CAD, Myocardial Perfusion SPECT, multi-label classification, transfer learning.**

I. Introduction

According to the World Health Organization (WHO), cardiovascular disease (CVD) is currently the leading cause of death globally, accounting for 32 percent of all deaths [1]. At the National Heart Conference 2017, a startling number was reported. Each year, Vietnam has about 200,000 people die from CVD, twice as many deaths from cancer. More importantly, the number of people suffering from CVD at a young age is increasing. Among CVD, coronary artery disease (CAD) and cerebral stroke are the leading causes of death or disability.

Coronary artery disease (CAD) [2] is the most common type of heart disease. It is sometimes called coronary heart disease or ischemic heart disease. CAD develops when the coronary arteries become too narrow, or cholesterol blockages (plaques) develop in the walls. Plaque consists of cholesterol, fatty substances, waste products, calcium and the clot-making substance fibrin. As plaque continues to collect on artery walls, arteries narrow and stiffen. Normally, there are three main obstructive regions of myocardium corresponding to three branches: left anterior descending artery (LAD), left circumflex artery (LCx), and right coronary artery (RCA) territories. This disease damages arteries and impedes supplying oxygen and blood to the heart. Eventually, the blood flow is reduced, causing chest pain (angina), shortness of breath, or other coronary artery disease signs and symptoms. A complete blockage can dangerously cause a heart attack. Particularly CAD caused by acute myocardial infarction, acute coronary syndrome can cause immediate death or lead to heart failure and death later. Moreover, diagnosing CAD usually requires many processes and experienced doctors. Human mistakes are sometimes unavoidable, and those flaws are dangerous, especially in clinical decision-making. Therefore, early and accurate detection of CAD becomes even more urgent nowadays.

In the light of technological developments, the abundance of modern machines are invented, which help diagnose diseases in general and CAD in particular. CAD can be detected by a combination of taking medical history with tests and imaging methods. Currently, there are several methods of diagnosing

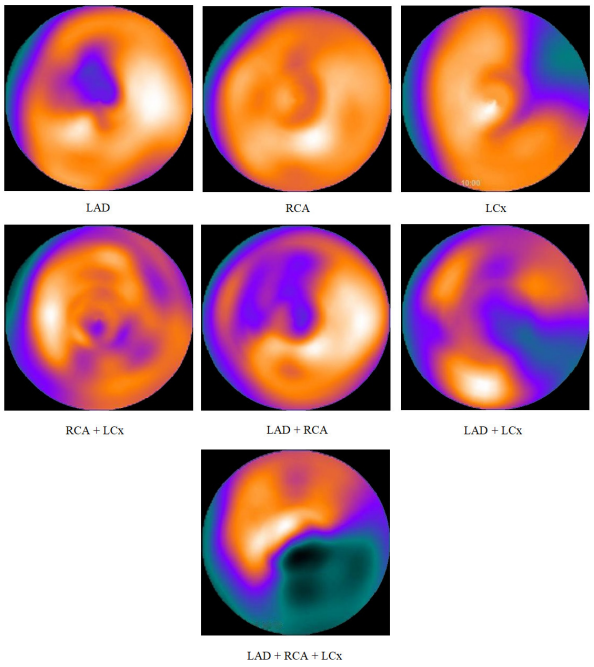

Fig. 1. Examples of SPECT images in our datasets.

CAD, such as electrocardiograph tests which record the electrical activity of the heart, a blood test that analyses factors in the blood that affect arteries, or computed tomography angiogram, which uses CT and contrast dye to view 3D pictures of moving heart and detect blockages in the coronary arteries. Among these techniques, Conventional single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) is one of the most widely used methods. SPECT MPI evaluates the present, extent, and degree of infarction by using gamma rays, providing 3-dimensional images. SPECT is based on the flow-dependent and/or metabolism-dependent selective uptake of a radioactive tracer by functional myocardial tissue. SPECT MPI has gained great success over the past decades as the modality of choice for accurately diagnosing patients with suspected coronary artery.

Many researches have been conducted [3]–[5], [9], [10], [12] based on SPECT MPI to help diagnose CAD accurately and reduce burden on doctors. In [12], four ensemble machine learning algorithms (Adaptive Boosting, Gradient Boosting, Random Forests, and Extreme Gradient Boosting) have been investigated in a dataset including one hundred and seven polar maps. Especially, the features extraction process generating inputs for four algorithms is intuitive and efficient. Each image was sliced into five horizontal and five vertical segments. After that, ten features were created by summing up pixel intensities from each segment. The results are impressive: all models had accuracy $>$ 90 percent and AUC approximately 0.8.

References [4] and [5] analyze 1638 (67% male) and 1160 (64% male) patients without known CAD, respectively. The authors apply deep learning models for polar maps and additional input images. Besides, sex information is also included to produce feature maps. Deep learning models are compared with current quantitative method (TPD - total perfusion deficit). The results demonstrate that deep learning models outperform TPD in terms of area under the receiver operating characteristic curve and sensitivity per patient and vessel.

TABLE I
STRESS POLAR MAPS CHARACTERISTICS.

Number	LAD	RCA	LCx	Image
Train	416	433	210	601
Test	102	109	53	200
Total	518	542	263	801

In [10], the authors utilize a predefined CNN-based model, termed RGB-CNN, which was proposed for other clinical problems, to solve binary-classification detecting CAD. RGB-CNN gives promising results (accuracy = 93.47% ± 2.81%, AUC score =0.936). The proposed methods are then compared with various state-of-the-art CNN backbones for the particular dataset.

Although many existing methods give impressive results in solving binary-classification to classify normal and abnormal SPECT images, none of them work on multi-label classification problems. That means, the computers now are able to predict very well whether or not a patient has disease, but having no clue about specific areas are being damaged.

The contribution of our research is two-fold:

1) We analyze the multi-label classification problem for SPECT images - which has not been studied before, in order to help further diagnose and give the patients and doctors information of areas causing CAD.
2) We propose a novel neural network-based structure solving multi-label classification problems in general.

The paper is organized as follows. In Section I introduces the process of generating datasets. Section III describes the proposed stack for classifying obstructive areas based on a fully connected neural network. Section IV presents the experimental evaluations and analysis. Finally, Section V gives the concluding remarks.

II. MATERIALS

This section generally presents the process of acquiring data. Our SPECT images datasets are collected at the Department of Nuclear Medicine, 108 Military Central Hospital from 2017 to 2019. The datasets were captured at stress by three kinds of specialized SPECT scanners (Infina, Optima, Ventri). It includes 801 polar maps, which are diagnosed with obstructive disease. Before the datasets were collected, all the patient's personal information was removed. This research was conducted with permission from the Department of Nuclear Medicine, 108 Military Central Hospital.

All images are read by three specialists with at least ten years of experience, trained in nuclear medicine in developed countries such as America, Japan, and Australia. Generating

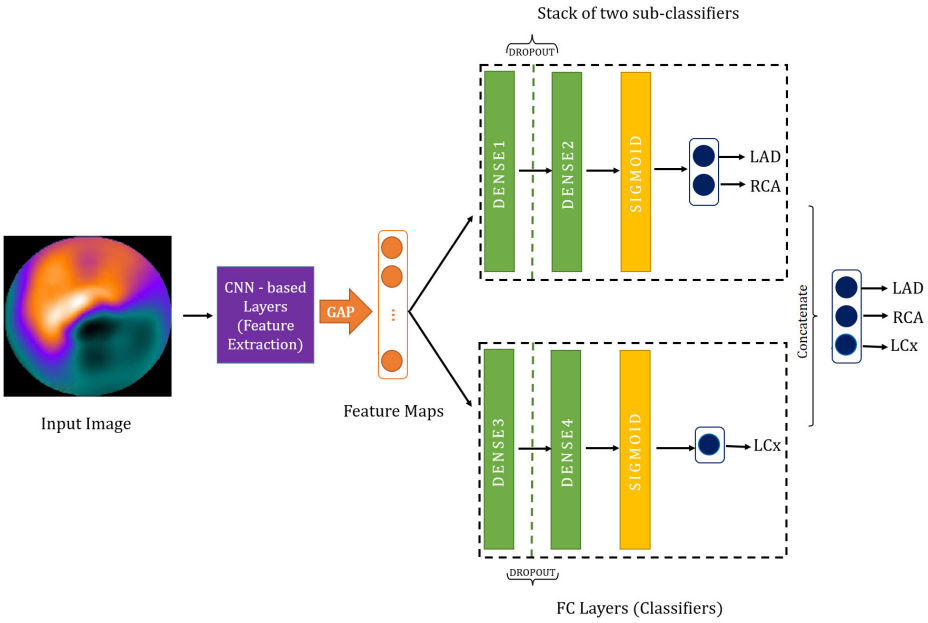

Fig. 2. An example of proposed stack of sub-classifiers solving multi-label classification problem.

SPECT images requires many processes and technical parts, from the pharmacist preparing the radiopharmaceuticals, the injection and imaging technicians, the image quality engineers, and doctors responsible for reading results. Every step requires high accuracy and correct sequence to ensure no errors in the whole implementation process.

A total of 1250 patients without known CAD, injected Tc^{99m_MIBI} (tracer). The patient's body weight determines rest-stress doses by a factor of 0.31mCi/kg. For patients unable to undergo stress physically, they get Diyridamole at a dose of 0,56mg/kg for 4 minutes after heart rate has reached 85 percent [8] [7], to increase the blood flow to the heart muscle as if patients were exercising.

After being injected, traces mix with blood and are taken up by the heart muscle as the blood flows through heart arteries. This radioactive material stored in the myocardium emits gamma rays with an energy level of 140 keV - which is captured by a special camera to show how well the heart muscle is perfused. We use Xeleris - a specialized program of Ge Healthcare for image reconstruction, processing 2D SPECT images, and integrating polar maps. After that, only stress polar maps are kept for further analysis.

Three specialists read and classified images as five levels from normal to surely abnormal. From these five categories, they are grouped more generally into two classes and each group is binary-labeled as normal or abnormal. In case of abnormal polar maps, obstructive areas are indicated, including the left anterior descending artery (LAD), the left circumflex artery (LCx), and the right coronary artery (RCA). From 1250 cases, merely 801 images with CAD and labeled obstructive areas are used to solve the multi-label classification problem. After pre-processing, RGB clinical images are exported in .png format with matrix size 352x352. Figure 1 above illustrates examples of all CAD cases having in our datasets. The datasets are separated into train and test set with the ratio 8:2 respectively. Because of the incidence of patients in Vietnam is uneven (usually LAD and RCA), our data are imbalanced. We can easily notice that the last label (LCx) merely has a half instances compared to others. Table I showcases our datasets in more details.

III. Proposed Stack of Classifiers

In this section, we describe in detail our proposed stack of sub-classifiers - a promising solution solving a multi-label classification problem for SPECT polar images.

For classification, after extracting useful features, classifiers solve the rest of problem. In most deep learning models, fully connected (FC) layers are the potential candidates to take responsibility for classifying objects. In [4], [5], [10], FC layers are also implemented to discern non-obstructive and obstructive SPECT images. However, for the multi-label classification problem, whether or not applying the same architecture to classify is good. Coming up with the idea of finding a multi-label classifier, which is less cumbersome but effective, we proposed a stack of sub-classifiers suitable for our datasets. The proposed idea can be applied to other multi-label classification problems.

Assume our multi-label classification has n labels, and we already found a good feature extractor that be able to extract useful image features. The intuitive idea is transforming multi-label classification to multi-binary classifications, by finding a suitable classifier for each label. Instead of using merely one classifier applied for all labels or a stack of n sub-classifiers for each label, we choose m ones ($0 < m < n$)

TABLE II
HYPERPARAMETERS OF NINE MODELS AFTER IMPLEMENTING HYPERBAND ALGORITHM.

Based model	Branch	Learning rate	Dense node		
			LAD	RCA	LCx
VGG16	3	0.01	576	640	192
	2	0.001	448		192
	1	0.01	64		
ResNet152V2	3	0.0001	640	128	576
	2	0.001	192		448
	1	0.001	128		
InceptionV3	3	0.0001	448	512	128
	2	0.0001	384		448
	1	0.001	640		

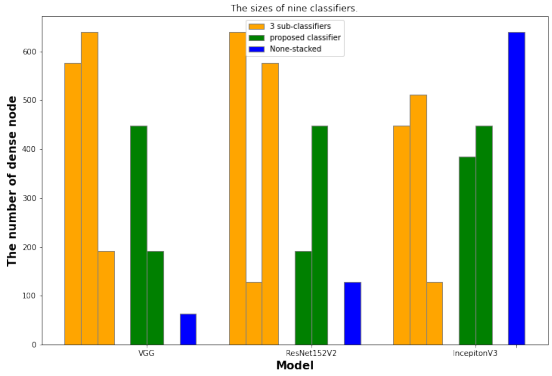

Fig. 3. A visual comparison of the number of dense nodes having in each model.

and allocate n labels into those sub-classifiers. The features extracted from the previous part are mutual-used as inputs of all sub-classifiers. Based on how balanced our data is and how well the sub-classifier can detect each label, we can find suitable m to construct our stack. For example, we can group easy-to-detect labels into one classifier and others with their own classifiers. In case of our datasets, we have n = 3 (LAD, RCA, LCx), and m = 2, the first two more easy-to-detect labels (LAD and RCA) are classified by the same branch. Figure 2 above illustrates our proposed stack of two sub-classifiers.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Implementation details

In this study, to evaluate the proposed method, we conduct totally nine experiments with three different approaches on three feature extractors utilized from VGG16 [11], ResNet152V2 [6], and InceptionV3 [13] models. We apply transfer learning and use pre-trained feature extractors with weights from Imagenet. Useful features are extracted with these based extractors. Parameters from that part of model are not updated through training process. We mainly focus on finding out suitable classifiers and assessing how well they perform on our datasets. For each based model, three kinds of classifier were analyzed for comparison: a non-stacked classifier, a stack of two, and three sub-classifiers. Totally, nine models are analyzed and evaluated (3 feature extractors x 3 classifiers). To find the optimal hyperparameters of each model, we use Keras Tuner library. Particularly Hyperband algorithm was chosen for all hypermodels.

In addition to the model architecture, we define hyperparameter search space for learning rate and the number of units in FC layers. Our search space has three learning rate (0.01, 0.001, 0.0001) and various dense node for each classifier (from 64 to 640, step 64). As a result, in the case of stacked classifiers, each branch is equivalent to one classifier, having its own hyperparameter. For example, search space in a stack of three branches is much larger than two sub-classifiers and a none-stacked model.

We optimize all networks rigidly with Adam algorithm, Binary cross-entropy for loss function, metrics using are Binary accuracy and area under the receive operating characteristics curve (AUC). For searching hyperparameters, we set objective = validation binary accuracy, max epochs = 100, and early stopping with patience = 5. After each FC layers, we use dropout = 0.2. Moreover, according to table I above, we can easily see that datasets are unbalanced. The number of images labeled LCx (263) is only half of the diagnosed LAD (518) and RCA (542). Therefore, we use different class weights for each label—specifically, LAD, RCA, and LCx with 1, 1, and 2, respectively. After using Hyperband algorithm, the hyperparameters of nine models are chosen and described in Table II and Figure 3.

TABLE III
EXPERIMENTAL RESULTS OF NINE DEEP LEARNING ARCHITECTURES FOR ALL LABELS.

Based model	Number of sub-classifier	Accuracy	AUC
VGG16	3	0.772	0.736
	2 (proposed method)	**0.787**	**0.755**
	1	0.722	0.656
ResNet152V2	3	0.749	0.710
	2 (proposed method)	**0.785**	**0.774**
	1	0.747	0.701
InceptionV3	3	0.741	0.662
	2 (proposed method)	**0.760**	**0.736**
	1	0.737	0.656

Nine deep learning models were implemented in Python 3 using open-source libraries, mainly are Tensorflow, Keras and Scikit-learn. All experiments are conducted on an HP computer equipped with an Intel Pentium (R) Core(TM) i5-8250U (1.60 GHz) and 8 GB of RAM, Windows 10 OS.

B. Results

After having hyperparameters, we train nine models with different epochs to optimize parameters with each architecture. We evaluate the model's performances by using the following metrics: accuracy, F1 score, recall, precision, and AUC. We evaluate those metrics for all labels and for each label separately. Table III and IV above showcase the performances of nine models in detail for all labels and each label, respectively.

TABLE IV
EXPERIMENTAL RESULTS OF NINE DEEP LEARNING ARCHITECTURES FOR EACH LABEL.

Based model	Area	Number of sub-classifier	Precision	Recall	F1-score
VGG16	LAD	3	**0.81**	0.84	**0.82**
		2(proposed method)	0.79	**0.85**	**0.82**
		1	0.79	0.76	0.77
	RCA	3	0.76	0.91	0.82
		2(proposed method)	**0.81**	0.89	**0.85**
		1	0.72	**0.97**	0.82
	LCx	3	0.71	**0.72**	**0.71**
		2(proposed method)	**0.74**	0.68	**0.71**
		1	0.68	0.37	0.48
ResNet152V2	LAD	3	0.81	0.77	**0.79**
		2(proposed method)	**0.82**	0.76	**0.79**
		1	0.81	**0.78**	**0.79**
	RCA	3	0.76	0.92	0.83
		2(proposed method)	**0.83**	0.86	**0.84**
		1	0.71	**0.96**	0.82
	LCx	3	0.67	0.60	0.63
		2(proposed method)	**0.73**	**0.75**	**0.74**
		1	0.67	0.68	0.68
InceptionV3	LAD	3	0.80	0.72	0.76
		2(proposed method)	**0.88**	**0.81**	**0.85**
		1	0.77	0.77	0.77
	RCA	3	0.82	**0.87**	**0.85**
		2(proposed method)	**0.85**	0.76	0.80
		1	0.84	0.84	0.84
	LCx	3	**0.80**	0.25	0.38
		2(proposed method)	0.64	**0.45**	**0.53**
		1	0.73	0.23	0.35

According to the results given in Table III above, our proposed method is outstanding in all three based models. In term of accuracy and AUC for all labels, our proposed approach completely outperforms the non-stacked and stack of three sub-classifiers models. In Table IV, performances of nine architectures are presented in particularly each label. In term of LCx - most difficult-to-detect label, our proposed method generally gives better results compared with other methods.

During experiments, we realize that in the case of three sub-classifiers - each label has its classifier, the model is fairly cumbersome. Moreover, the results even drop while having more parameters. Specifically, after just a few epochs, the value of loss function of model during the training process does not improve when the number of epochs increases. The model encountered a vanishing gradient problem.

In terms of non-stacked architecture, the model is not much different from traditional binary classification. Instead of having two units, the last FC layer now comprise n units, where n is the number of labels in multi-label classification. For example, in our datasets, n = 3. This kind of model can work well in binary classification, the idea can be found in [4], [5], [10]. However, as our results have proved, most non-stacked models have the poorest performance because of trouble to predict LCx. In our experiments, the models can detect with much higher accuracy for the first two labels (LAD, RCA) compared to the last label (LCx). The reason is that our dataset is imbalanced and detecting LCx is harder than the other two labels because the number of images having this label (263) is only half of LAD (518) and RCA (542) (Table I). Non-stacked architecture may still detect well on balanced datasets. However, imbalanced labels in multi-label problems are sometimes unavoidable, especially when the number of labels increases significantly.

The remaining problems of the two models above are addressed in terms of the two-sub-classifier model. By finding suitable branches for our stack and allocating labels in the appropriate sub-classifiers, we can deal with cumbersome architecture and give better result on unbalanced datasets. The results in Table III demonstrate the efficiency of our proposed method. This stack of two sub-classifiers effectively solves the multi-label classification problem with promising results.

V. CONCLUSION

This paper applied transfer learning methods, utilizing pre-trained features extractors from pre-knowledge of Imagenet, and proposed a practical stack of classifiers solving multi-label classification problems. We conduct extensive experiments on different models and classifiers on clinical polar SPECT datasets. Experimental results demonstrate that our proposed stack works well and gives the most outstanding results when combined with various features extractors while keeping a moderate number of parameters.

More importantly, our method can give better results in imbalanced datasets for each label - one of the most challenging obstacles that multi-label classification commonly faces. Our proposed method can help solve other multi-label classification problems with promising efficiency and accuracy. In the future, we will work more to address multi-label classification for imbalanced SPECT images datasets and go further with others

multi-label classification problems. Hopefully, our study is able to apply in solving other multi-label classification problems for diverse kinds of objects and datasets.

REFERENCES

[1] Cardiovascular diseases. World Health Organization, https://www.who.int/health-topics/cardiovascular-diseases, accessed on 2022-06-14

[2] Coronary artery disease. Mayo Foundation for Medical Education and Research (May 2022), https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613, accessed on 2022-06-14

[3] Apostolopoulos, I., Papathanasiou, N., Spyridonidis, T., Apostolopoulos, D.: Automatic characterization of myocardial perfusion imaging polar maps employing deep learning and data augmentation. Hellenic journal of nuclear medicine **23** (07 2020). https://doi.org/10.1967/s002449912101

[4] Betancur, J., Commandeur, F., Motlagh, M., Sharir, T., Einstein, A.J., Bokhari, S., Fish, M.B., Ruddy, T.D., Kaufmann, P., Sinusas, A.J., et al.: Deep learning for prediction of obstructive disease from fast myocardial perfusion spect: a multicenter study. JACC: Cardiovascular Imaging **11**(11), 1654–1663 (2018)

[5] Betancur, J., Hu, L.H., Commandeur, F., Sharir, T., Einstein, A.J., Fish, M.B., Ruddy, T.D., Kaufmann, P.A., Sinusas, A.J., Miller, E.J., et al.: Deep learning analysis of upright-supine high-efficiency spect myocardial perfusion imaging for prediction of obstructive coronary artery disease: a multicenter study. Journal of Nuclear Medicine **60**(5), 664–670 (2019)

[6] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. pp. 770–778 (06 2016). https://doi.org/10.1109/CVPR.2016.90

[7] Hesse, B., Tägil, K., Cuocolo, A., Anagnostopoulos, C., Bardiès, M., Bax, J., Bengel, F., Busemann Sokole, E., Davies, G., Dondi, M., et al.: Eanm/esc procedural guidelines for myocardial perfusion imaging in nuclear cardiology. European journal of nuclear medicine and molecular imaging **32**(7), 855–897 (2005)

[8] Holly, T., Abbott, B., Al-Mallah, M., Calnon, D., Cohen, M., DiFilippo, F., Ficaro, E., Freeman, M., Hendel, R., Jain, D., Leonard, S., Nichols, K., Polk, D., Soman, P.: Single photon-emission computed tomography (10 2010). https://doi.org/10.1007/s12350-010-9246-y

[9] Kaplan Berkaya, S., Ak, I., Gunal, S.: Classification models for spect myocardial perfusion imaging. Computers in Biology and Medicine **123**, 103893 (07 2020). https://doi.org/10.1016/j.compbiomed.2020.103893

[10] Papandrianos, N., Papageorgiou, E.: Automatic diagnosis of coronary artery disease in spect myocardial perfusion imaging employing deep learning. Applied Sciences **11**(14), 6362 (2021)

[11] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556 (09 2014)

[12] de Souza Filho, E.M., Fernandes, F.d.A., Wiefels, C., de Carvalho, L.N.D., dos Santos, T.F., dos Santos, A.A.S.M.D., Mesquita, E.T., Seixas, F.L., Chow, B.J.W., Mesquita, C.T., Gismondi, R.A.: Machine learning algorithms to distinguish myocardial perfusion spect polar maps. Frontiers in Cardiovascular Medicine **8**, 1437 (2021). https://doi.org/10.3389/fcvm.2021.741667, https://www.frontiersin.org/article/10.3389/fcvm.2021.741667

[13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. pp. 2818–2826 (06 2016). https://doi.org/10.1109/CVPR.2016.308