Detection of Atherosclerosis Using Deep Learning

Juliet Chebet Moso[1,2,*], Mohamed Tahar Bennai[1], Zahia Guessoum[1], Stéphane Cormier[1] and Jasminka Hasić Telalović[3]

[1]*Lab-I* UR 4474, Université de Reims Champagne-Ardenne, 51100, Reims, France*
[2]*Department of Computer Science, Dedan Kimathi University of Technology, Private bag 10143, Nyeri, Kenya*
[3]*University Sarajevo School of Science and Technology, 71000, Sarajevo, Bosnia and Herzegovina*

Abstract

Atherosclerosis is one of the major causes of cardiovascular disease, requiring early detection and intervention. Deep learning techniques, particularly transfer learning, offer potential avenues for improving the diagnosis and management of atherosclerosis. In this paper, a transfer learning approach is proposed to enhance the detection of atherosclerotic plaque by adapting pre-trained Convolutional Neural Networks (CNNs). By fine-tuning these models on a dataset of medical images, the study aims to leverage learned representations to improve detection accuracy and efficiency. Additionally, data augmentation is used to enhance model robustness and address data scarcity and class imbalance issues. The findings from our experiments indicate that the ResNet-50 model outperformed others in terms of Recall at 1.0, followed by Inception-v3 at 0.941. In classification accuracy, ResNet-50 achieved 93%, followed by the Inception-v3 model at 86%. Similarly, in AUC-ROC performance, the ResNet-50 model attained the highest score of 0.99, with the Inception-v3 model following closely at 0.966. Our results demonstrate that transfer learning significantly improves the accuracy, sensitivity, and specificity of the detection of atherosclerotic plaque, showcasing its potential as a valuable tool for the detection of atherosclerosis.

Keywords

Deep learning, Image classification, Medical Imaging, Transfer learning

1. Introduction

Atherosclerosis is the most prevalent underlying cause of cardiovascular disease (CVD) [1], a worldwide health concern that affects the heart and circulatory system [2]. It stands as the leading cause of global mortality, responsible for 17.9 million deaths in 2019 alone, according to WHO [3]. In European Society of Cardiology member countries, CVD prevails as the primary cause of mortality, affecting more females than males, with ischemic heart disease constituting a significant portion of fatalities [4]. In cardiovascular medicine, Coronary artery calcification (CAC) serves as a crucial clinical marker indicating atherosclerosis progression characterized by the accumulation of calcium deposits in the coronary arteries.

RIF'24: The 13th International Conference on Research in computIng at Feminine, May 20–21, 2024, Constantine, Algeria
*Corresponding author.
juliet-chebet.moso@univ-reims.fr (J. C. Moso); m.bennai@univ-boumerdes.dz (M. T. Bennai); zahia.guessoum@univ-reims.fr (Z. Guessoum); stephane.cormier@univ-reims.fr (S. Cormier); jasminka.hasic@ssst.edu.ba (J. H. Telalović)
0000-0002-0276-0164 (J. C. Moso); 0000-0002-7364-5171 (M. T. Bennai); 0000-0003-1366-8889 (Z. Guessoum); 0000-0003-4507-4815 (S. Cormier); 0000-0002-8540-5052 (J. H. Telalović)

Detecting CAC is crucial for assessing coronary heart disease risk and guiding preventive treatments. Tests like treadmill tests, radionuclide scans, CT scans, MRI scans, and coronary angiography aid in CAC identification [5]. Early and precise detection enables timely interventions, alleviating the burden of cardiovascular disease. Manual interpretation of medical imaging for CAC diagnosis is labour-intensive and prone to inter-observer variability [6], necessitating more efficient and reliable diagnostic methods. Over the past decade, medical research has made noteworthy progress using Convolutional Neural Networks (CNNs) for image processing. Several approaches have been proposed for detecting and classifying CAC from CT scans, including paired CNN for CAC quantification and risk classification [7], Recurrent CNN for plaque and stenosis classification [8], and 3D CNN architecture for atherosclerosis visualization [9].

Deep learning holds promise for transforming CAC detection, automating procedures and enhancing diagnostic accuracy. However, its integration into cardiac CT is an evolving process due to the need for in-depth validation [10]. The scarcity of labelled medical data for training these systems also poses a challenge, as models suffer from overfitting and increased complexity. Optimizing hyperparameters such as filter size and learning rate poses challenges. Generalization errors occur when processing images from varied modalities or dealing with unseen pathological cases, underscoring the need for robust deep learning models in CAC detection [11]. A viable method for producing highly accurate classification models with little training data is transfer learning [6]. It leverages a network that has undergone comprehensive prior training on a specific dataset and reconfigures its network components to meet the demands of the new domain.

This study utilizes CNNs to automatically classify CT scans for detecting presence of atherosclerosis. It highlights the effectiveness of transfer learning in leveraging existing knowledge and resources to address the challenges associated with detection of atherosclerotic plaques in medical imaging. By harnessing the power of deep learning and transfer learning techniques, we aim to contribute to the development of more efficient and accurate diagnostic tools for cardiovascular diseases, ultimately leading to improved patient outcomes. This will improve detection performance and alleviate cardiologist workload. Three pretrained models: ResNet-50, Inception V3, and VGG19 where implemented and compared against our proposed baseline CNN in classification of images into diseased or normal categories. This technique simplifies clinical processes, enhances early identification of at-risk patients, allows personalized risk assessment, and facilitates prompt interventions.

This paper is structured as follows: Section 2 presents the related works on classification of coronary artery diseases. Section 3 presents the methodology which discusses the data acquisition and pre-processing steps and our proposed classification models leveraging transfer learning. Section 4 is dedicated to the experimental results, with section 5 giving the conclusion and perspectives.

2. Related Work

Computer vision research on coronary artery diseases is difficult because of resource constraints brought on by the laws governing patient privacy. The difficulty is increased by the intricacy and accuracy needed for these investigations. This section provides an overview of existing research in our context, shedding light on relevant studies and works.

Deep learning methods, including artificial neural networks and CNNs, have been used to detect coronary artery diseases. These architectures eliminate the need for explicit feature engineering [12], as they autonomously extract relevant information from training data reducing computational complexity. In [7], a paired CNN technique was applied to quantify coronary artery disease (CAD) on CT angiography scans, achieving a sensitivity of 0.71 and an accuracy of 83% compared to manual annotations by an expert human observer. A Recurrent CNN was used for detection of coronary artery plaque and stenosis in CT scans, achieving 77% accuracy for plaque analysis and 80% for stenosis analysis [8]. Candemir et al. [9] developed a 3D CNN architecture using visual cues to categorize vessels, provide insights into coronary artery volume, identify pathological lesions, and automatically pinpoint atherosclerosis regions, achieving an accuracy of 90.9%.

Gupta et al. [13] investigated transforming coronary CT images from 3D to 2D, focusing on straightened-Multiplanar Reformatted (MPR) representations for arteriosclerosis prediction. Using Inception-v3 with transfer learning and data augmentation, they achieved an AUC of 0.93. Alothman et al. [14] proposed a method that employed feature extraction and a CNN model. Their modified DenseNet-161 with transfer learning and leaky ReLU activation achieving 99.2% prediction accuracy and F1 score of 0.9895, and precision-recall curves of 0.92 and 0.91 with less memory usage and computation time compared to state-of-the-art CNNs.

In [15], a CAD detection method using You Only Look Once (YOLO) V7 and UNet++ models is proposed. A fuzzy function is used to enhance images and extract key features. The Aquila optimization algorithm is used to optimize hyperparameters in the UNet++ model. This approach reduces computational costs and improves the model's performance, achieving an average accuracy of 99.40% and an AUROC of 0.97. In [16], non-contrast and contrast heart CT scan images were utilized to predict stenosis. Transfer learning with five pretrained neural network models was applied, with EfficientNetB0 achieving the best recall of 0.933. In[5], three CNN models (Inception ResNet v2, VGG, and ResNet 50) were trained for coronary artery calcification detection. ResNet 50 achieved the highest accuracy of 98.52% on cardiac cropped images, obtained after dissecting the cardiac region using K-means clustering and mathematical morphology.

This section reviewed works showcasing efficient coronary artery disease detection using deep learning. Deep learning methods excel in detecting detailed features and handling noise to a certain extent during data processing. However, they face limitations such as insufficient training data, imbalanced datasets leading to biased models, and challenges with overfitting and underfitting, which can be mitigated through hyperparameter tuning. To address these issues, we developed a transfer learning-based deep learning method to detect atherosclerosis using minimal data.

3. Methodology

In this section, we present the procedures for data acquisition and pre-processing. Additionally, we delve into our proposed classification models, which utilize transfer learning, along with the

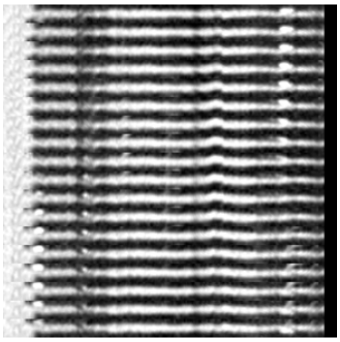

Figure 1: Representation of a diseased coronary artery with atherosclerosis.

Table 1
Data Description

	Train set	Validation set	Test set
Diseased	2364	50	175
Normal	2304	50	1116

evaluation metrics used to assess their performance.

3.1. Data acquisition and Pre-processing

This study uses publicly available Coronary Computed Tomography Angiography (CCTA) images from 500 patients [17] which is de-identified and anonymized to protect privacy and confidentiality. The images underwent a 2D Projection Reformatting procedure. The coronary artery was projected from 18 different angles, generating a 2D projection for each angle at 10-degree intervals, to boost accuracy and maintain a realistic representation [13]. A 2D representation of a diseased coronary artery projected from 18 angles is shown in Figure 1.

The dataset was partitioned into training, validation, and testing subsets on a per-patient basis, following a ratio of 3:1:1 (300/100/100). This ensures a fair proportion of both healthy, atherosclerosis-free cases and sick instances, each making up 50% of the total. To improve modelling and attain dataset balance, data augmentation was done to artery images obtained from the 300 training cases, resulting in a six-fold increase in the dataset to 2,364 images. This augmentation aimed to improve model training and maintain dataset balance. Markedly, the entire validation dataset, the testing dataset, and the normal component of the training dataset (2,304 images) were excluded from augmentation. Table 1 provides a summary of the pre-processed dataset:

3.2. Transfer Learning Methodology

Three pre-trained models, VGG19 [18], Inception-v3 [19], and ResNet50 [20], were selected to be used for classification using transfer learning. Inception-v3 excels in handling extensive datasets and diverse image dimensions and resolutions, proving particularly beneficial in medical

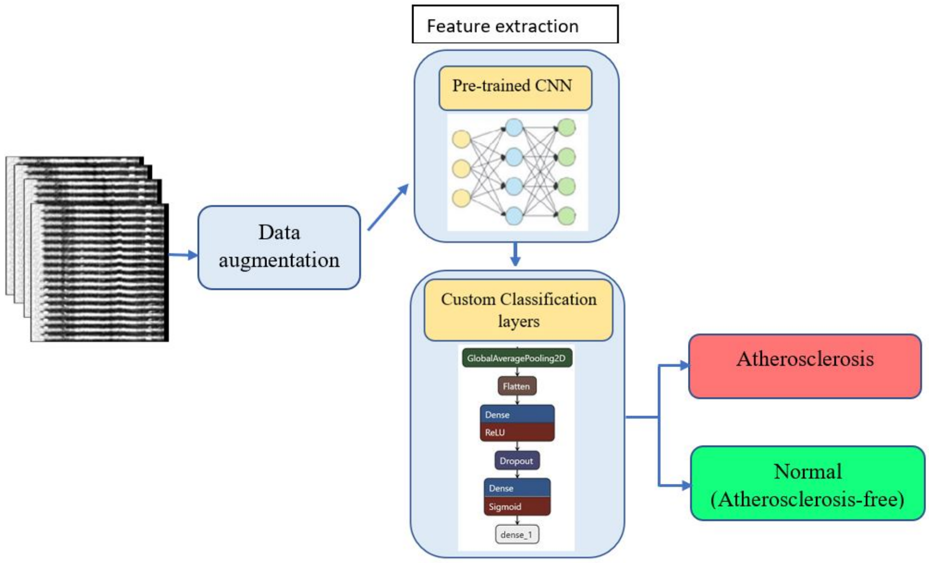

Figure 2: Overview of the transfer learning based workflow applied in detection of atherosclerosis from CCTA images

imaging. ResNet50 introduces residual connections, enabling the network to learn residual functions for mapping input to output. These connections tackle the vanishing gradient problem by acting as gradient superhighways, ensuring uninterrupted gradient propagation in deeper architectures.

Transfer learning with the pre-trained models involved freezing all layers of the models and removing their Fully Connected layers. Global Average Pooling 2D (GAP) was then applied to the frozen layers to extract features for subsequent fine-tuning layers. GAP effectively reduces spatial dimensions while preserving channel information, aiding in parameter reduction and guarding against overfitting. By condensing each feature map into a single value, GAP maintains spatial information, enhancing model generalization and adaptability to new datasets. For the binary classification task, the sigmoid activation function was chosen, compressing inputs to a range of 0 to 1. Outputs near 1 indicate a high probability of being "Diseased" while those close to 0 suggest "Normal" affiliation. Binary Cross-Entropy Loss, paired with sigmoid activation, effectively measures dissimilarity between predicted probabilities and true labels. The transfer learning based workflow applying pre-trained deep learning model for feature extraction and the custom classification layers for detection of atherosclerosis is summarized in Figure 2.

Proposed Baseline CNN: While deep transfer learning models may demonstrate state-of-the-art performance on certain datasets and tasks, it's essential to compare their performance against simpler baseline approaches to assess whether the additional complexity and computational cost are justified. Our proposed baseline CNN serves as a fundamental benchmark against which the efficacy of deep transfer learning models can be evaluated. Our baseline CNN is designed to offer insights into the effectiveness of these sophisticated approaches.

The architecture of our baseline CNN begins with pre-processed image data fed into a

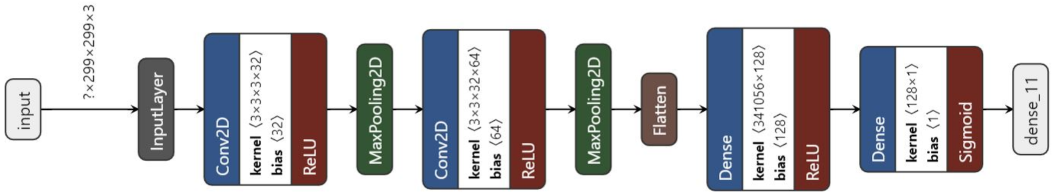

Figure 3: Proposed Baseline CNN Network Architecture

Convolutional Layer (Conv2D) with 32 filters, facilitating feature detection. Rectified Linear Unit (ReLU) activation introduces non-linearity, enabling complex pattern learning. Subsequent Max Pooling layer downsamples feature maps, reducing spatial dimensions and computational load while retaining essential information. Another Conv2D layer with 64 filters and ReLU activation is employed, followed by further Max Pooling. The model then flattens the 2D feature maps into a 1D vector, feeding it into a Fully Connected Layer (Dense) comprising 128 neurons with ReLU activation. Finally, a single neuron with sigmoid activation is utilized for binary classification, employing binary cross-entropy loss and the Adam optimizer. Figure 3 give a visual representation of our proposed baseline CNN's architecture.

3.3. Performance Indicators

The effectiveness of the proposed model was evaluated using Recall, Precision, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and accuracy. These metrics are derived from the following measures [21]:

- True positive (TP): Accurately identified images with atherosclerosis.
- False positive (FP): Incorrectly classified normal images.
- True negative (TN): Correctly classified normal atherosclerosis-free images.
- False negative (FN): Images with atherosclerosis that are incorrectly classified as normal images.

Recall or Sensitivity evaluates the model's ability to correctly identify all relevant instances within a dataset. Specificity measures the proportion of correctly predicted negative instances out of all the actual negative instances. The positive prediction rate is represented by precision, and the classification performance is evaluated in terms of both recall and precision using the F-score. Accuracy measures the model's total classification performance, accounting for both True Positives and True Negatives. In classification tasks, the AUC-ROC curve is used as a performance statistic across various threshold settings [22], with higher AUC-ROC values indicating superior detection capabilities.

4. Implementation and Experiments

In this section, details about the implementation of the networks using transfer learning process will be discussed alongside the results achieved through tests.

Table 2
Performance evaluation of the models on test dataset

	Precision	Recall	Specificity	F1 Score	AUC	Accuracy
ResNet-50	0.548	1.0	0.923	0.708	0.990	0.93
Inception-v3	0.372	0.941	0.852	0.533	0.966	0.860
VGG19	0.165	0.765	0.639	0.271	0.813	0.650
Baseline CNN	0.181	0.765	0.678	0.292	0.836	0.685
UNet++ [15]	0.985	0.986	-	0.986	0.97	0.994
PCCN [14]	0.982	0.985	0.987	0.983	-	0.989

4.1. Model Implementation

The model scripts were developed using the **Keras** framework, utilizing **Tensorflow** as the backend within a Python 3 Jupyter notebook environment. The experiments were conducted on an NVIDIA P100 GPU equipped with 16 GB of high-bandwidth memory. Initialization of the ResNet-50, Inception v3, and VGG19 networks involved using weights pre-trained on the **ImageNet** dataset. During the training phase, batches of 32 images were utilized, evenly divided between positive and negative cases of atherosclerosis. After 50 epochs of training, the best model was saved for further analysis and evaluation of the test dataset.

4.2. Results

After training the models, an evaluation was conducted using a distinct test dataset to gauge their performance. Table 2 presents a summary of the results obtained from this test dataset, consisting of a sample size of 200 images. Comprehensive evaluation of models in disease classification tasks is crucial due to the potentially significant consequences of false positives and false negatives. In the diagnosis of atherosclerosis, false negatives can have severe consequences; ensuring that all positive cases are correctly identified is crucial. False negatives in medical diagnosis can lead to missed diagnoses, delayed treatments, and potentially serious health consequences for patients. Recall measures the ability of a model to capture all positive cases out of the total number of actual positives. In scenarios where correctly identifying true negatives is essential, Specificity is applied. Specificity indicates the model's ability to correctly identify individuals without a particular condition as being negative for that condition, minimizing false alarms or misdiagnoses.

In our experiments, Recall is given more weight since it aims to minimize false negatives [16], as it focuses on capturing as many true positive cases as possible. The confusion matrix plays a pivotal role in this assessment by offering an in-depth analysis of the models' predictions and how well they match the ground truth. Figure 4 illustrates the results of the confusion matrix for the four models. According to the findings, ResNet-50 successfully classified all diseased images correctly. In contrast, Inception-v3 misclassified one instance in which disease was present as absent (FN), with four FNs for both VGG19 and baseline CNN. Regarding false positives, where the models incorrectly predicted the presence of disease when it was not present, ResNet-50 had the lowest number at 14, whereas VGG19 had the highest number at 66.

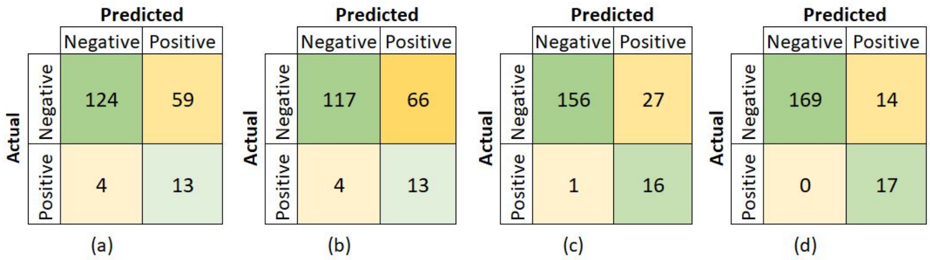

Figure 4: Performance of the models on the test dataset: (a) Baseline CNN, (b) VGG19, (c) Inception-v3, (d) ResNet-50

Class imbalances are typical in medical image classification when there are much fewer positive instances (i.e., ill patients) than negative cases (i.e., healthy patients). This is the situation in this instance, where there are 17 positive cases out of 200 cases. AUC-ROC is robust to class imbalance and provides an aggregated performance measure that is not influenced by the class distribution. This makes it particularly suitable for evaluating our classifiers since the positive cases (diseased instances) are less prevalent. This metric offers a singular, succinct value that effectively quantifies the model's performance across different thresholds, considering both sensitivity (true positive rate) and specificity (true negative rate). In medical image classification, where the balance between sensitivity and specificity is crucial, AUC-ROC offers a holistic view of the classifier's ability to distinguish between classes. As depicted in Figure 5, it is evident that ResNet-50 exhibits the highest discrimination capacity with an AUC-ROC of 0.99, followed by Inception-v3 at 0.966, baseline CNN at 0.836, and VGG19 at 0.813. When compared to other state-of-the-art models, UNet++ [15] and PCCN [14], ResNet-50 had the best performance in both Recall and AUC-ROC.

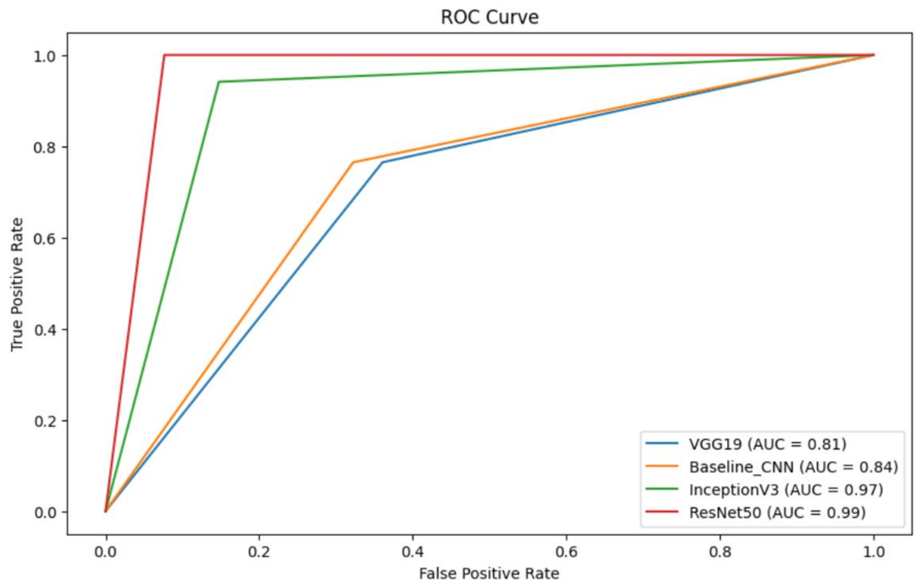

Figure 5: AUC-ROC curve indicating the performance of the models on the test dataset

5. Conclusion

In this study, we proposed a deep learning approach to analyse heart CT scans for detecting atherosclerosis. We employed a transfer learning strategy, utilizing three pre-trained deep neural networks: ResNet-50, VGG19, and Inception v3. These models were compared against a baseline shallow CNN model to assess their performance. Our proposed baseline CNN gives comparable results to the more complex models outperforming the VGG19 model. The ResNet-50 network exhibited promising results, achieving a Recall of 1.0, an accuracy of 93% and an AUC-ROC of 0.99. This approach shows potential for effective classification of atherosclerosis in CT scans.

Future work will include an evaluation of the model's robustness across different imaging modalities, acquisition settings, and patient demographics (such as different races, ethnicities, genders, ages, and socioeconomic backgrounds) to ensure its generalizability and applicability in diverse clinical environments. Interpretability and explainability of our deep learning models require further attention so as to understand the decision-making process. Additionally, clinical validation is crucial to ascertain the real-world applicability of our approach. While our results are promising, further validation in clinical settings is necessary to assess the model's performance in routine practice.

6. Acknowledgement

This work was made possible by funding from the European Union's Horizon Europe program for widening participation and spreading excellence under Grant Agreement number 101060145 for EDIRE project.

Declaration on Generative AI

While preparing this work, the authors used Grammarly to check grammar and spelling. After using this tool, they reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] N. D. Wong, M. J. Budoff, K. Ferdinand, I. M. Graham, E. D. Michos, T. Reddy, M. D. Shapiro, P. P. Toth, Atherosclerotic cardiovascular disease risk assessment: an american society for preventive cardiology clinical practice statement, American journal of preventive cardiology 10 (2022) 100335.

[2] N. J. Pagidipati, T. A. Gaziano, Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement, Circulation 127 (2013) 749–756.

[3] W. H. Organization, Cardiovascular diseases (cvds), 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[4] A. Timmis, P. Vardas, N. Townsend, A. Torbica, H. Katus, D. De Smedt, C. P. Gale, A. P. Maggioni, S. E. Petersen, R. Huculeci, et al., European society of cardiology: cardiovascular disease statistics 2021, European Heart Journal 43 (2022) 716–799.
[5] S. Lee, B. Rim, S.-S. Jou, H.-W. Gil, X. Jia, A. Lee, M. Hong, Deep-learning-based coronary artery calcium detection from ct image, Sensors 21 (2021) 7059.
[6] P. Kora, C. P. Ooi, O. Faust, U. Raghavendra, A. Gudigar, W. Y. Chan, K. Meenakshi, K. Swaraja, P. Plawiak, U. R. Acharya, Transfer learning techniques for medical image analysis: A review, Biocybernetics and Biomedical Engineering 42 (2022) 79–107.
[7] J. M. Wolterink, T. Leiner, B. D. de Vos, R. W. van Hamersvelt, M. A. Viergever, I. Išgum, Automatic coronary artery calcium scoring in cardiac ct angiography using paired convolutional neural networks, Medical image analysis 34 (2016) 123–136.
[8] M. Zreik, R. W. Van Hamersvelt, J. M. Wolterink, T. Leiner, M. A. Viergever, I. Išgum, A recurrent cnn for automatic detection and classification of coronary artery plaque and stenosis in coronary ct angiography, IEEE transactions on medical imaging 38 (2018) 1588–1598.
[9] S. Candemir, R. D. White, M. Demirer, V. Gupta, M. T. Bigelow, L. M. Prevedello, B. S. Erdal, Automated coronary artery atherosclerosis detection and weakly supervised localization on coronary ct angiography with a deep 3-dimensional convolutional neural network, Computerized Medical Imaging and Graphics 83 (2020) 101721.
[10] C. B. Monti, M. Codari, M. van Assen, C. N. De Cecco, R. Vliegenthart, Machine learning and deep neural networks applications in computed tomography for coronary artery disease and myocardial perfusion, Journal of thoracic imaging 35 (2020) S58–S65.
[11] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, IEEE journal of biomedical and health informatics 21 (2016) 4–21.
[12] M. D. Zeiler, R. Fergus, Visualizing and understanding convolutional networks, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13, Springer International Publishing, Cham, 2014, pp. 818–833.
[13] V. Gupta, M. Demirer, M. Bigelow, K. J. Little, S. Candemir, L. M. Prevedello, R. D. White, T. P. O'Donnell, M. Wels, B. S. Erdal, Performance of a deep neural network algorithm based on a small medical image dataset: incremental impact of 3d-to-2d reformation combined with novel data augmentation, photometric conversion, or transfer learning, Journal of digital imaging 33 (2020) 431–438.
[14] A. F. AlOthman, A. R. W. Sait, T. A. Alhussain, Detecting coronary artery disease from computed tomography images using a deep learning technique, Diagnostics 12 (2022) 2073.
[15] A. R. Wahab Sait, A. K. Dutta, Developing a deep-learning-based coronary artery disease detection technique using computer tomography images, Diagnostics 13 (2023) 1312.
[16] M. Aono, T. Asakawa, H. Shinoda, K. Shimizu, T. Komoda, Predicting stenosis in coronary arteries based on deep neural network using non-contrast and contrast cardiac ct images, in: Proceedings of the 2023 6th International Conference on Machine Vision and Applications, 2023, pp. 154–160.
[17] M. Demirer, V. Gupta, M. Bigelow, B. Erdal, L. Prevedello, R. White, Image dataset for a

cnn algorithm development to detect coronary atherosclerosis in coronary ct angiography, Mendeley Data 1 (2019). URL: https://data.mendeley.com/datasets/fk6rys63h9/1. doi:10.17632/fk6rys63h9.1.

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

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 2818–2826. doi:10.1109/CVPR.2016.308.

[20] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[21] G. Canbek, S. Sagiroglu, T. T. Temizel, N. Baykal, Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights, in: 2017 International Conference on Computer Science and Engineering (UBMK), IEEE, 2017, pp. 821–826.

[22] G. O. Campos, A. Zimek, J. Sander, R. J. Campello, B. Micenková, E. Schubert, I. Assent, M. E. Houle, On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study, Data mining and knowledge discovery 30 (2016) 891–927.