Journal of Electrical Engineering and Computer (JEECOM)
Vol. xx, No. xx (20xx), DOI: 10.33650/jeecom.v4i2
p-ISSN: 2715-0410; e-ISSN: 2715-6427

Improving Computer-Aided Medical Diagnosis Using Generative Adversarial Networks for Carotid Artery Ultrasound Image Data Augmentation and Classification

João Gonçalves[1], Ricardo Fitas [2]
[1] Faculty of Engineering, University of Porto, R. Dr. Roberto Frias s/n, 4200-465 Porto, Portugal
[2] Technical University of Darmstadt, 64289 Darmstadt, Germany

Article Info

Article history:

Diterima xx Bulan 20xx
Revisi xx Bulan 20xx
Diterbitkan xx Bulan, 20xx

Keywords:

CNN
GAN
Medical Images
Carotid Artery

ABSTRACT

Cardiovascular diseases are a leading cause of death globally, making early detection of atherosclerosis critical for prevention. Carotid artery ultrasound imaging is a common diagnostic tool; however, the limited availability of labelled medical images hinders the training of deep learning models. This study examines generative adversarial networks (GANs) for data augmentation and classification of carotid artery Doppler images to improve computer-aided medical diagnosis. Four convolutional neural networks (CNNs) – AlexNet, VGGNet, GoogleNet, and CifarNet – are evaluated for their classification performance on original and extended datasets. AlexNet outperforms the other models, achieving a classification accuracy of 94.18% on the extended dataset. The GAN implementation for data augmentation and overfitting reduction demonstrates the potential of generative models in enhancing the performance of deep learning models in medical image analysis, particularly in the "common artery carotid" class. This research contributes to understanding GANs as a valuable tool for data augmentation and classification in the context of carotid artery ultrasound images.

Corresponding Author:
Ricardo Fitas,
Technical University of Darmstadt, 64289 Darmstadt, Germany
Email: ricardo.fitas@tu-darmstadt.de

1. INTRODUCTION

Cardiovascular diseases are among the major causes of death in the world [1]. An early symptom is atherosclerosis [2], [3] and its early detection is crucial for prevention. Ultrasound imaging of the carotid artery is commonly performed for such prevention. It is characterized in the literature as a popular, affordable and non-intrusive method [4], [5].

The growth of deep learning neural networks knowledge and applications have been exponential and critical in almost all aspects of modern society, such as security, social media, industry optimization, economy and healthcare. Also, taking advantage of such advancements, several studies were conducted, e.g. [6]–[9], to segment the carotid artery automatically. It would allow a faster measurement of atherosclerosis risk, because analyzing such ultrasound images is very difficult for medical staff [5]. According to the literature, the risk of disease can be predicted using parameters that can be extracted from image analysis, such as artery stiffness, lumen diameter, and thickness of the arteries [5], [10]. Therefore, automatic segmentation provides a promising approach for aiding prevention based on pattern recognition, data analysis and control [11].

One of the recurrent challenges of training deep models is overcoming overfitting. When the dataset is scarce, patterns and regularities on training images have reasonably higher accuracy than unprocessed ones and classification may not be correct. Such transformations can be geometric, involve colour alteration, kernel filters and feature space augmentation [12], [13] and must preserve the given label [14].

Another challenge is the insufficient availability of medical images required to optimize the number of parameters of deep neural networks to maintain their accuracy. Also, collecting and labelling images individually is costly and can compromise patients' personal data. Common solutions to increase the original image dataset include data augmentation and data generation techniques. In [15], the authors explore the use of generative adversarial networks (GANs) for data augmentation in Computed Tomography (CT) image segmentation tasks. Specifically, the authors used a variant of GANs called CycleGAN to generate synthetic CT images and augment a limited amount of labelled CT data. They evaluated the effectiveness of CycleGAN-generated synthetic images and found that they captured the range of variability in authentic CT images. Training deep learning models on both the original and synthetic images demonstrated that the augmented dataset could significantly improve segmentation performance. In [16], the authors investigate the performance of deep convolutional generative adversarial networks (DCGANs) in synthesizing ultrasound images. The authors empirically analyze various DCGAN architectures and training strategies to improve image quality and realism.

As verified in a recent literature review [17], there is still a need for more research in areas such as image generation to address the limited availability of labelled data. The authors suggest that data augmentation and generative models such as GANs could be used to increase the size and diversity of training datasets of medical images and that further research is needed in this area. They also highlight the importance of addressing issues such as interpretability and generalizability in deep learning models for medical image analysis. After an extensive literature review, the authors did not find a paper to implement generated data representative of carotid arteries.

This paper focuses on appropriating a neural network classifier for the recognition of the carotid artery Doppler transversal images, aided by data synthesis via developed GAN, aiming at tackling such difficulties. Therefore, the most appropriate generative model to fit our dataset is studied.

The following Research Questions are expected to be discussed and answered:
RQ1. Are classification accuracy results feasible enough to prove that a neural network approach could be used for a clinical coronary artery diagnosis?
RQ2. Is the implementation of a GAN positive for data augmentation and consequent reduction of model overfitting?
RQ3. What conclusions are achieved when comparing the performance from selected state-of-the-art classification models?

2. RELATED WORKS

The section begins by highlighting the importance of automated methods to support the diagnosis of deformations in medical images, especially with the application of neural network models for the classification and synthesis of new images for deeper training. It then discusses various related works in the field, specifically for diagnosing coronary artery disease and carotid artery disease.

In [18], the authors implemented a Generative Adversarial framework for spatial feature extraction and fusion of medical and surgical state images. This allowed them to achieve real-time path planning for endovascular surgeries. The study used an extensive dataset of 350000 samples validated by experienced surgeons. The results showed the great application of real-time path planning for robotic and novice training in endovascular surgery.

Similarly, [16] used a DCGAN approach to expand their dataset of ultrasound images of the carotid artery to overcome the problem of costly annotated samples and anonymized data. The expanded dataset was then used to train a better classifier for detecting and preventing early stages of rare diseases.

Another related work [19] proposed a new GAN approach called Anatomical GAN to support other automated coronary artery disease diagnosis methods. The model specializes in correcting anatomically invalid pixel-wise segmentation on longitudinal ultrasound images. The model has the potential to automate the examination of the shape and texture of artery walls, which requires an expert and is critical to saving lives and minimizing costs. However, one important limitation mentioned is misdiagnosing defects due to speckle noise in the image and the faulty quality distribution, reflection, and absorption of ultrasound waves by dense structures.
In [20], the authors compared the performance of classification convolutional neural networks (CNNs) trained from scratch, pre-trained "off the shelf" models, and unsupervised pre-trained CNNs in the context of chest X-

rays and CT lung nodule identification. The models assessed include a seven-layer "AlexNet-CNN," [14] a shallower "CifarNet," [21], [22] and a variant of the "GoogleNet-CNN" [23]. The lung nodule (LN) patches used in the study had a resolution of 64x64 pixels, which were up-sampled via bi-linear interpolation to 256x256 pixels to fit the input dimensions of the AlexNet and GoogleNet models.

The study found that CifarNet was undertrained due to the limited input resolution. At the same time, AlexNet had inferior performance, indicating that pre-trained ImageNet CNNs alone may not be optimal as off-the-shelf deep image feature extractors. On the other hand, GoogleNet was susceptible to overfitting and performed poorly in the study. The authors suggested that transfer learning offers a much better initialization of CNN parameters than random initialization, boosting the performance of deep learning models.

In [24], an unsupervised GAN was used to identify abnormal regions of interest in clinical images. The study focused on benchmarking with lung nodules in CT scans, microcalcification in X-ray mammography images for breast screening, and macular fluid in OCT scans of the retina. A significant problem studied is the limited prediction power of annotated training data due to being limited to known markers. A new aspect of the GAN is then introduced, which can detect unprecedented anomalies that do not fit the distribution and complement the vocabulary of clinical markers.

Finally, [25] experimented with deep learning classification methodologies using 2D echocardiogram and 3D Doppler images. The main goal was to detect and classify valvular heart flow patterns using colour imaging. The authors used Variational Auto Encoder (VAE) with Convolutional Neural Network (CNN) for feature extraction and space reduction. The study demonstrated the potential of VAE-CNN in detecting and classifying valvular heart flow patterns. However, the authors did not compare their results with other state-of-the-art models.

The related works demonstrate the importance of using deep learning and generative models to address the limitations of limited annotated data for medical image diagnosis. While each study had different goals, datasets, and methodologies, deep learning and generative models show promise for improving the accuracy and efficiency of medical image diagnosis. However, it is essential to carefully consider the limitations and potential biases of the models and datasets used in the studies to avoid misdiagnosis and improve patient outcomes.

3. BACKGROUND OF CONCEPTS

3.1. Neural Networks

In broader terms, neural network systems use training examples to infer rules for detecting a certain target. A conventional network has an input, output, and hidden layers composed of several neurons. Some hidden layers work as embeddings, i.e., low-dimensional spaces where we can relate items from high-dimensional vectors so that similar inputs end up together. It is a crucial resource of our architecture since it can segregate images with similar parameters. To help achieve the desired output, a cost function will quantify how well the bias and weights from the output come close to the desired result, and by minimizing this function, the optimal weight-bias definition is sorted out. Due to high dimensional input, the vectorial cost function C has many variables. Finding the global minimum requires defining the gradient of C, a vector of partial derivatives, and applying the stochastic gradient descent technique. It estimates the gradient function from a small sample of randomly chosen training inputs (mini-batch), followed by picking out another randomly chosen batch and training with those until the training data is exhausted and an epoch is completed.

3.2. Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are a deep learning algorithm commonly used in image and video recognition tasks. They are specifically designed to process data with a grid-like topology, such as images, by employing a mathematical operation called convolution.
In a typical CNN architecture, the input image is first passed through convolutional layers, followed by non-linear activation functions and pooling layers. The convolutional layers learn local patterns or features in the image, which are combined and abstracted by subsequent layers to form higher-level representations.
A cross-entropy cost function is used as the loss or objective function to train a CNN. This function measures the difference between the predicted output and the true label and aims to minimize this difference.

Examples of CNN well-established in the literature are CifarNet, VGGNet, AlexNet and GoogleNet. CifarNet is a convolutional network variant of the CIFAR-10 dataset (6000 images for every 10 classes, with 32x32 pixels) model definition [22], [26], prominent in the literature in [20], [26]. VGGNet is a deeper network, compared to CifarNet, that had great success in a few generations of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [20], [27]. AlexNet was published in [14] excelled over other learning methods for ILSVRC 2012. It has five convolution layers, three pooling layers, and two fully connected layers with approximately 60 million free parameters. GoogleNet, introduced in [23], is elaborated deeper than other CNN architectures. Initially designed for the fixed image dimension of 256x256x3 pixels, it innovates with introducing the "Inception layer", which merges filters of different dimensions into one. The model has two convolution layers, two pooling layers and nine "Inception" layers. More details about the estimated number of parameters are listed in Table 1.

Table 1. Number of parameters of different models

Network	AlexNet	CifarNet	GoogleNet	VGGNet
Number of Parameters (in millions)	60	1.1	6.8	144

3.3. Generative Adversarial Network (GAN)

A Generative Adversarial Network (GAN) consists of a "two-player min-max game" [28] between a generator and a discriminative model. At the same time, the generator's goal is to learn the given data distribution from a random noise vector input, which in our case are the medical images; a discriminative model determines whether the generated sample is a synthetic or real data distribution.

CNNs and GANs are both types of deep learning algorithms but have different purposes and architectures. While CNNs are mainly used for image and video recognition tasks and employ convolutional layers to learn local patterns and features in the input data, GANs are generative models that learn the underlying distribution of the training data and generate new samples that are similar to the real ones.

Both models are trained with iterative methods using highly efficient dropout algorithms [14] and backpropagation. To optimize the adversarial process, maximization of the discriminative probability of labeling the distribution given correctly is performed. At the same time, the generator's goal is to minimize a function, increasing the adversarial model failure probability. Finding this equilibrium with high dimensional parameters and non-convex cost functions is challenging and may lead to convergence errors [29], and, for this reason, in practice, it is used a gradient descent technique to find the minimum of a cost function. In detail, the Adam optimizer was chosen based on adaptive estimation of first and second-order moments. According to [30], it is *"computationally efficient, has little memory requirement, […], and is well suited for problems that are large in terms of data/parameters".*

Minibatch discrimination is a technique used to improve GAN training [31], by preventing the generator from collapsing to a state where it emits the same image. Since the discriminator processes each sample separately, there is no coordination between the gradients. When they point to similar directions for many images, the generator's output will be constant. In this way, a correct distribution of parameters cannot be achieved as the point which generator considers to be a reality is being directed by discriminative gradients around the space forever. Generated and real samples are fed to a minibatch layer to correct this problem, which returns image features that facilitate the discriminator task to classify single examples.

Training the CNN and GAN discriminative model for classification and detection of medical images Region Of Interest (ROI) is made using supervised training, which consists in feeding the model labelled input and minimizing the cross entropy between the observed labels and model predictive distribution. It is important to mention that the manual labelling of training data refers to image captions and not a pixel-wise segmentation of the ROI. This topic is further explored in the next section.

4. METHODOLOGY

ISSN: 2715-6427

In this section, the dataset characteristics and appropriate processing are explored in detail, various CNN architectures are introduced, and their performance is compared. Figure 1 shows the flowchart for the proposed methodology. As shown in it, the present methodology starts by preparing the original dataset for CNN implementation, followed by training the proposed classification networks and electing the most adequate for classification. In a second iteration, after applying post-processing filters to GAN-obtained images and extending our dataset, the studied CNN models are trained, and the optimal one is chosen for classification.

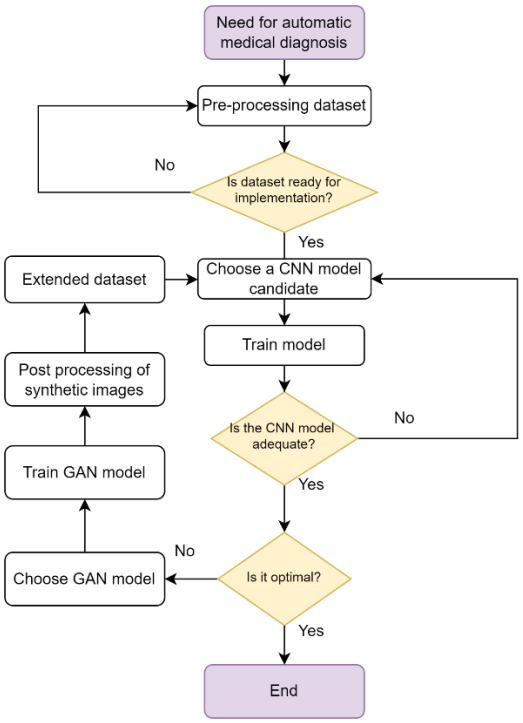

Figure 1. Flowchart of the implemented methodology

4.1. Datasets

A database of 21 Doppler 2D videos of the transversal view of carotid arteries is used. The bifurcation of one terminal of a common carotid artery, Figure 2a, to form internal and external carotid arteries, Figure 2b [32], is evident in all videos, as a single common artery slowly transitioning to two distinct arteries can be recognized on most of them, and the reverse process on others.

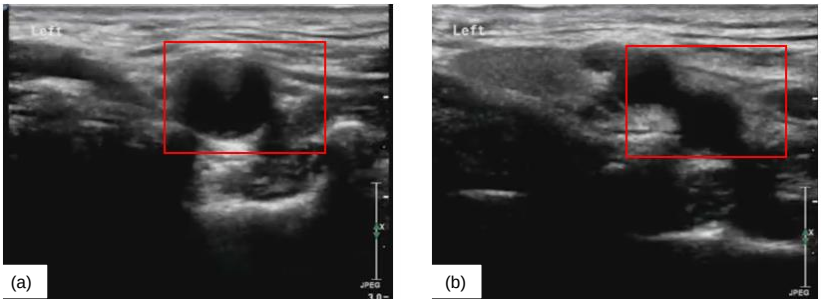

Figure 2. Side-by-side comparison of two frames. The left image shows the common carotid artery and the right one shows the internal and external carotid artery. With a red square indicating the estimated region of transition.

4.2. Frame selection

To implement the deep learning neural networks, and since the models were chosen to use images as inputs, 6419 frames are extracted from the mentioned videos, which incorporate the training and validation dataset. This process was made using Python programming language and OpenCV library. The resolution of these images is 600 x 450 pixels. Through a manual selection process, we can segregate them into 3116 frames for the common carotid artery and 3303 frames for internal and external carotid arteries after bifurcation. We use the same training and validation set for all CNN models studied for better benchmarking.

4.3. Pre-processing

Following the data-centric approach [33], the performance of the studied CNN models is enhanced by improving the quality of training data and minimizing the impacts of using a relatively small dataset. For this, inadequate examples from the training batch are removed, such as those seen in Figure 3, to show unambiguous examples for each class and clarify the network labelling instructions. It only applies to the training batch since the validation batch is preferred to be hyper-realistic.

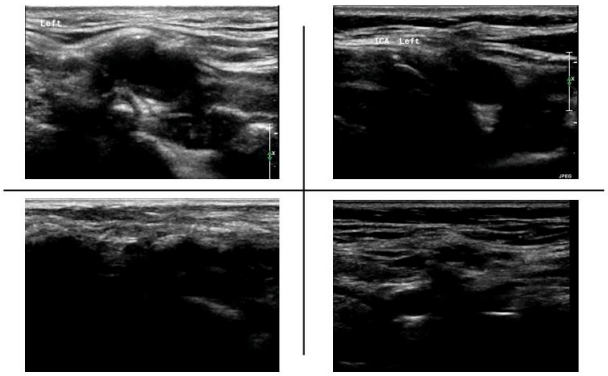

Figure 3. Examples of deleted images due to poor quality

Due to the high precision of CNN-based image processing, accurate artery segmentation is unnecessary, unlike in [18], in which real-time path planning demands a precise separation from other human tissues. The localization of regions of interest within the image is learned through the selectively weighted neurons in the deepest convolutional layers during training [34], [35]. CNN learns to ignore the peripheral noise of the images, setting small filter weights to those regions [36].

4.4. CNN architectures

The convolutional neural network architectures chosen for study are AlexNet, GoogleNet, CifarNet and VGGNet. They are tailored to accommodate an input of 284x284x3 pixels. Current literature has supported the authors' choice of using high-parameter networks [20], while other works rely on smaller CNN architectures with fewer parameters, such as [26], [37].

Keras is a deep learning API written in Python, developed to enable fast experimentation, running on top of the machine learning platform Tensorflow. Relying on the keras application from TensorFlow (v2.11.0), a few lines of code returned a dataset with batches of images from two subdirectories: class "common artery" and class "internal and external carotid arteries". Each batch has 32 images with a resolution of 284x284 pixels. This resolution is achieved through bilinear interpolation. Furthermore, it automatically shuffles images with a random seed and breaks the data into two fractions, 20% are reserved for validation and 80% for training.

For each one of the architectures applied, the parameters for an input of 284x284 image resolution are tailored. While in CifarNet, we increased the filter size, in AlexNet, GoogleNet and VGGNet, which are deeper networks, the filter size, stride and pooling parameters were reduced, aiming at achieving promising results. Despite being "simplified" versions of the original architectures, they are suited for smaller training datasets.

4.5. CNN training

CNN models are learnt from scratch; all the parameters are initialized with random Gaussian distributions and trained for 10 epochs with a mini-batch size of 32 images. A Binary Cross Entropy is used as a loss function

and an Adam optimizer with a default learning rate of 0.001. The authors have used an NVIDIA GeForce RTX 3060 GPU for model training.

4.6. GAN Architecture

Two different GAN models are used. The discriminator and generator layer's parameters are shown in greater detail in Figure 4.

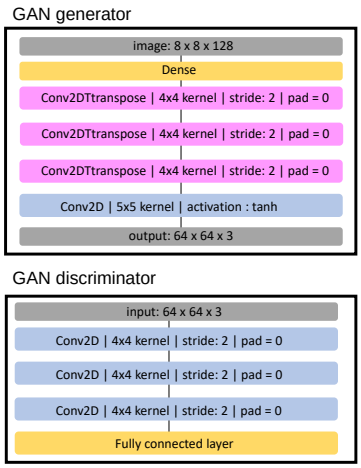

Figure 4. GAN architecture

4.7. GAN training

Generator and discriminator networks are trained using stochastic gradient descent with a mini-batch size of 32 images. Pre-processing consists of loading images and resizing them to 64x64 pixels, followed by a conversion to a numpy array and scaling is applied to the input images to get the range of -1 to 1 for tanh activation.

4.8. Post-processing of GAN-obtained images

Data quality may not be optimal during image acquisition and formation due to noise, an unwanted component from the random variation of brightness and colours. An anisotropic diffusion [38] is used to enhance and denoise the training data since it is more adequate than a Median filter followed by a Gaussian filter, which is proven to be an effective method for smoothing and better edge estimation [39]–[41]. After this, we applied a 3x3 sharpen filter [42] and a bilinear interpolation to rescale back to an image resolution of 284x284 pixels. Results from GAN trained model extended our dataset in 714 "common artery" frames and 600 "internal and external carotid arteries" frames, being that these numbers followed a manual selection from a data-centric perspective [33]. A sample of generated images can be seen in Figure 5.

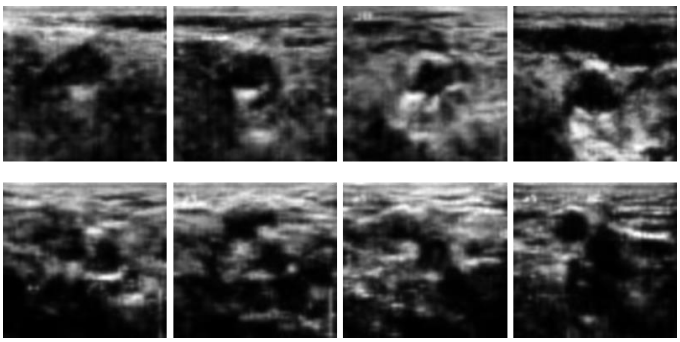

Figure 5. Sample of generated images. Upper images: common carotid artery; Lower images: external and internal arteries

5. RESULTS AND DISCUSSION

5.1. CNN training results

Results from the common and external/internal arteries original dataset classification are reported in Table 2. AlexNet yielded the most competitive classification accuracy values of the four investigated CNN models, while VGGNet performed below expectations. The performance of VGGNet might be due to overfitting since millions of parameters are required to train such a deeper network, and the optimal training dataset might be extensively bigger and more diverse than ours. Similarly, GoogleNet falls short compared to AlexNet and CifarNet, in terms of classification accuracy, and available data samples were insufficient to train this deep network properly from random initialization. CifarNet training accuracy being higher than others is unusual, considering that, as in [20], the network would be expected to be under-trained due to the narrow input resolution and parameter complexity; however, the validation accuracy fits better in our expectations.

Table 2. Common, external and internal arteries detection results using various CNN models. Numbers in bold indicate the best classification accuracy values (original dataset)

	Training		**Validation**	
	Loss	**Accuracy**	**Loss**	**Accuracy**
AlexNet	0.1248	**0.9241**	0.1305	**0.9244**
GoogleNet	0.137	0.9176	0.1698	0.9072
CifarNet	0.261	**0.9268**	0.2397	0.9096
VGGNet	0.6931	0.5149	0.6924	0.5199

Comparing CNN performances from the original dataset with the extended dataset, reported in Table 3, it is evident that increasing training data had a positive result on the accuracy of the models, successfully preventing overfitting. Similar to previous analysis, CifarNet accuracy results are unexpected since it is as high as AlexNet and GoogleNet scores. The fact that a shallower network obtained a similar training accuracy to deeper networks may suggest that the training dataset may be insufficient in quantity and quality. Averse to the general tendency, the VGGNet accuracy score decreased slightly for both training and validation sets. Being the deepest network, data augmentation was expected to have less effect on preventing overfitting. One can argue that integration of synthetic images was the cause of the score decline. Such a complex model tries to fit the given data too closely and tends to capture noisy images instigated by anomalous or unexpected features on synthetic data. Analogous to the original dataset, AlexNet obtained the higher validation accuracy for the extended dataset, which classifies it as the optimal classification network. In Figures 6 and 7, in terms of accuracy results, it is clear that during training, the accuracy score quickly got high and stayed somewhat constant. During validation, accuracy scores wobble around training values tending to stay higher with the increase of epochs. This behaviour matched the expectations and resembled a good deep-learning training graph. In terms of loss, for both graphs with epoch increase, the training loss score had a steady decrease, while the validation loss score had an unsteady reduction.

Table 3. Common, external and internal arteries detection results using various CNN models. Numbers in bold indicate the best classification accuracy values (extended dataset)

	Training		**Validation**	
	Loss	**Accuracy**	**Loss**	**Accuracy**
AlexNet	0.1139	0.9324	0.1136	**0.9418**
GoogleNet	0.4082	0.9279	0.3382	0.9334
CifarNet	0.2079	0.9358	0.1897	0.9366
VGGNet	0.6930	0.5120	0.6933	0.4987

ISSN: 2715-6427

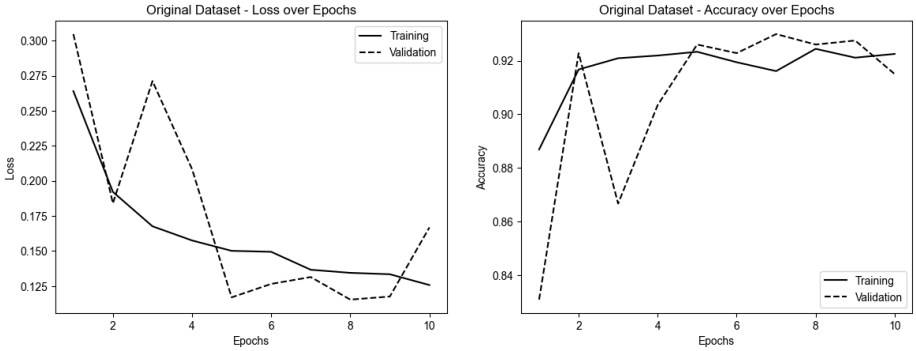

Figure 6: Training and validation for both loss and accuracy during training of AlexNet without the addition of artificial images (original dataset)

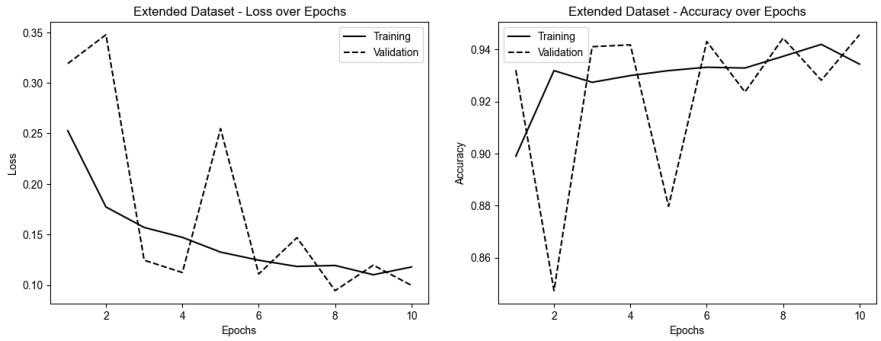

Figure 7: Training and validation for both loss and accuracy during training of AlexNet with the addition of artificial images (extended dataset)

Validation accuracy results are generally lower and less steady than training values since models tend to fit the training closely, resulting in high accuracy and low performance on unseen data. As explained in the "Pre-processing" section, we removed ambiguous examples from the training batch.

5.2. Confusion matrix

A confusion matrix resorts to measuring the performance of the GAN model since it is a good representation of multiclass classification problems. The resulting confusion matrix can be seen in Figure 8. The trained AlexNet model was used to classify the synthetic images as it showed the best accuracy compared to the other CNNs. The True Positive ("TP") value shows the fraction of correctly predicted labels given by AlexNet, from the batch of synthetic single common artery images. Conversely, the True Negative ("TN") value represents the fraction of correctly predicted labels from AlexNet from the generated batch of internal and external carotid arteries. Such a high TP value means the GAN model captures and replicates common artery images with high precision. However, close values of TN and False Negative ("FN") show that the GAN model performed poorly in replicating with detail internal and carotid arteries. Despite being able to distinguish the two different classes of images in the synthetic batch, the results show that the generative model could not find the correct distribution of image parameters, which could be hard for human eyes to perceive, which is crucial for AlexNet classification. In other words, since the supervised learning method has not restrained the networks to a pixel-wise analysis, the medical image ROIs that the generative model gives more importance might not be the same as those considered by the classification model, which caused the discrepancies between classes in the confusion matrix.

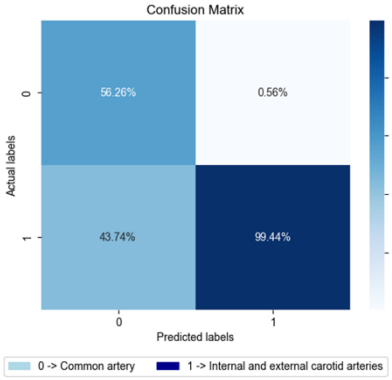

Figure 8. Confusion matrix for multiclass classification in percentage

5.3. Discussion of the research questions

RQ1. Are classification accuracy results feasible enough to prove that a neural network approach could be used for a clinical coronary artery diagnosis?

Experimental results aim for a viable CNN application in real-time clinical coronary artery diagnosis. Despite the high classification accuracy, this conclusion does not imply that the proposed CNN models are adequate for medical operations. Further testing is required to secure data accuracy and reliability. Furthermore, besides regulatory compliance, machine learning implementation depends on user acceptance, integration with existing systems, backup and disaster recovery.

RQ2. Is the implementation of a GAN positive for data augmentation and consequent reduction of model overfitting?

Implementing a GAN model combined with post-processing filtering resulted in positive data augmentation for "common artery carotid" dataset. Concerning the "internal and external carotid arteries" dataset, when observing class discrepancies on the confusion matrix, we concluded the generative model could not replicate the given data as authentic as before. Generally, complex CNN models performed better on unseen data and were agile in capturing the ROI instead of noise in training data, indicating a reduction of model overfitting.

RQ3. What conclusions are achieved when comparing the performance from selected state-of-the-art classification models?

When comparing results from the studied classification models, apart from concluding that AlexNet outperformed other models, it is important to recognize that CifarNet, a simpler network, achieved better scores than deeper, extended models, such as GoogleNet and VGGNet, which indicates that, despite reducing overfitting, data augmentation lacked in diversity, quantity and quality to bring out the full potential from deeper models.

6. CONCLUSIONS AND FUTURE WORKS

Machine learning applications can identify patterns and be an important tool in diagnosis and triage by analysing medical data, imaging scans, and patient history. Generating unique, unseen data from training examples to extend the dataset for a classification model is a familiar method in the literature since it is crucial in reducing costs and time when collecting samples needed to prevent the overfitting of deep learning models. The hypothesis was to settle for the fittest model in a pool of classification CNN models and tailor a generative model to obtain realistic variants for two classes of carotid artery Doppler images, common and external/internal, to extend our training dataset to improve classification accuracy. After extensive scientific review, the authors did not find a paper with similar objectives in this specific context to complement and believe that a gap was filled in the literature.

With the original dataset of 6419 medical images, AlexNet achieved the highest classification scores with 92,41% training accuracy and 92,44% validation accuracy. After extending our dataset with 714 "common artery" frames and 600 "internal and external carotid arteries" frames, AlexNet continued to be the leading

candidate in classification with 93.24% training and 94,18% validation accuracy. Results comply with the hypothesis, yet overfitting is still a predominant limitation in our experiment, and in future works, data diversity, quantity and quality must be reviewed and improved. Image transformations must be explored since, similar to lymph nodes classified in [20], carotid arteries have no pre-determined anatomic orientation and applying geometric transformations that may alter orientation, such as reflection, rotation and translation, is a natural choice for data augmentation. In terms of performance metrics, the chosen GAN model excelled in generating artificial "common artery carotid" images, with 99,44% of predicted labels being correct. In comparison, it returned unclear artificial "internal and external carotid arteries" with 56,26% correct predicted labels.

Besides increasing training data, the authors argue that improving noise regularization and the complexity of post-processing image treatment are key steps to improve the generative model in further studies.

REFERENCES

[1] I. Poledník, J. Sulzenko, and P. Widimsky, "Risk of a coronary event in patients after ischemic stroke or transient ischemic attack," Anatol. J. Cardiol., vol. 25, no. 3, pp. 152–155, 2021, doi: 10.5152/AnatolJCardiol.2021.75548.
[2] G. A. Roth et al., "Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015.," J. Am. Coll. Cardiol., vol. 70, no. 1, pp. 1–25, Jul. 2017, doi: 10.1016/j.jacc.2017.04.052.
[3] S. Fu, F. Yan, and Y. Long, "A new automatic segmentation algorithm for carotid artery ultrasound images," in 2022 International Seminar on Computer Science and Engineering Technology (SCSET), 2022, pp. 110–113. doi: 10.1109/SCSET55041.2022.00033.
[4] E. Ukwatta, D. Buchanan, G. Parraga, and A. Fenster, "Three-dimensional Ultrasound Imaging of Carotid Atherosclerosis," 2011 Int. Conf. Intell. Comput. Bio-Medical Instrum., pp. 81–84, 2011.
[5] J. H. Gagan et al., "Automated Segmentation of Common Carotid Artery in Ultrasound Images," IEEE Access, vol. 10, pp. 58419–58430, 2022, doi: 10.1109/ACCESS.2022.3179402.
[6] R.-M. Menchón-Lara and J.-L. Sancho-Gómez, "Fully automatic segmentation of ultrasound common carotid artery images based on machine learning," Neurocomputing, vol. 151, pp. 161–167, 2015, doi: https://doi.org/10.1016/j.neucom.2014.09.066.
[7] D. Samiappan and V. Chakrapani, "Classification of carotid artery abnormalities in ultrasound images using an artificial neural classifier," Int. Arab J. Inf. Technol., vol. 13, pp. 756–762, 2016.
[8] C. F. Castro, R. Fitas, and E. Azevedo, "AUTOMATIC SEGMENTATION IN TRANSVERSE ULTRASOUND B- MODE IMAGES OF THE CAROTID ARTERY," vol. 30, pp. 71–76, 2018.
[9] R. Fitas, C. Castro, L. Sousa, C. António, R. Santos, and E. Azevedo, "Analysis of Sequential Transverse B-Mode Ultrasound Images of the Carotid Artery Bifurcation," 2019, pp. 521–530. doi: 10.1007/978-3-030-32040-9_53.
[10] A. Anand and N. R. Gurram, "Automated Deep Learning-based Single-Step Diameter Estimation of Carotid Arteries in B-mode Ultrasound.," Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf., vol. 2022, pp. 434–437, Jul. 2022, doi: 10.1109/EMBC48229.2022.9871254.
[11] C. M. Bishop, "Neural networks and their applications," Rev. Sci. Instrum., vol. 65, no. 6, pp. 1803–1832, 1994, doi: 10.1063/1.1144830.
[12] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges," J. Digit. Imaging, vol. 32, no. 4, pp. 582–596, 2019, doi: 10.1007/s10278-019-00227-x.
[13] S. Jamroziński and U. Markowska-Kaczmar, "Semi-supervised classifier guided by discriminator," Sci. Rep., vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-18947-6.
[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Curran Associates Inc., 2012, pp. 1097–1105.
[15] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," Sci. Rep., vol. 9, no. 1, p. 16884, 2019, doi: 10.1038/s41598-019-52737-x.
[16] D. Kumar, M. A. Mehta, and I. Chatterjee, "Empirical Analysis of Deep Convolutional Generative Adversarial Network for Ultrasound Image Synthesis," Open Biomed. Eng. J., vol. 15, pp. 71 – 77, 2021, doi: 10.2174/1874120702115010071.
[17] G. Litjens et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.
[18] Y. Zhao et al., "Surgical GAN: Towards real-time path planning for passive flexible tools in endovascular surgeries," Neurocomputing, vol. 500, pp. 567 – 580, 2022, doi: 10.1016/j.neucom.2022.05.044.
[19] M. Engin et al., "AGAN: An Anatomy Corrector Conditional Generative Adversarial Network," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12262 LNCS, pp. 708 – 717, 2020, doi: 10.1007/978-3-030-59713-9_68.
[20] H.-C. Shin et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Trans. Med. Imaging, vol. 35, pp. 1285–1298, 2016.
[21] H. R. Roth et al., "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1170–1181, 2016, doi: 10.1109/TMI.2015.2482920.
[22] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," 2009.
[23] C. Szegedy et al., "Going Deeper With Convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[24] C. Park, S. Lim, D. Cha, and J. Jeong, "Fv-AD: F-AnoGAN Based Anomaly Detection in Chromate Process for Smart Manufacturing," Appl. Sci., vol. 12, no. 15, 2022, doi: 10.3390/app12157549.
[25] I. Wahlang et al., "Deep learning methods for classification of certain abnormalities in echocardiography," Electron., vol. 10, no. 4, pp. 1 – 20, 2021, doi: 10.3390/electronics10040495.
[26] H. Roth et al., "Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation," IEEE Trans. Med. Imaging, vol. 35, 2015, doi: 10.1109/TMI.2015.2482920.
[27] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR, vol. abs/1409.1,

2014.
[28] I. Goodfellow et al., “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
[29] I. Goodfellow, “On distinguishability criteria for estimating generative models,” 2014.
[30] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” CoRR, vol. abs/1412.6, 2014.
[31] T. Salimans et al., “Improved Techniques for Training GANs,” in Advances in Neural Information Processing Systems, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., Curran Associates, Inc., 2016. [Online]. Available: https://proceedings.neurips.cc/paper/2016/file/8a3363abe792db2d8761d6403605aeb7-Paper.pdf
[32] C. Hacking and J. Jones, “Internal carotid artery,” Radiopaedia.org. Radiopaedia.org, 2008. doi: 10.53347/rID-4524.
[33] O. H. Hamid, “From Model-Centric to Data-Centric AI: A Paradigm Shift or Rather a Complementary Approach?,” in 2022 8th International Conference on Information Technology Trends (ITT), 2022, pp. 196–199. doi: 10.1109/ITT56123.2022.9863935.
[34] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Is object localization for free? - Weakly-supervised learning with convolutional neural networks,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 685–694. doi: 10.1109/CVPR.2015.7298668.
[35] M. A. Nielsen, “Neural_Networks_and_Deep_Learning”.
[36] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1717–1724. doi: 10.1109/CVPR.2014.222.
[37] J. Shin, N. Tajbakhsh, R. T. Hurst, C. Kendall, and J. Liang, “Automating Carotid Intima-Media Thickness Video Interpretation with Convolutional Neural Networks,” 2017.
[38] C. P. Behrenbruch, S. Petroudi, S. Bond, J. D. Declerck, F. J. Leong, and J. M. Brady, “Image filtering techniques for medical image post-processing: an overview,” Br. J. Radiol., vol. 77, no. suppl_2, pp. S126–S132, 2004, doi: 10.1259/bjr/17464219.
[39] A. Kumar and S. S. Sodhi, “Comparative Analysis of Gaussian Filter, Median Filter and Denoise Autoenocoder,” in 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), 2020, pp. 45–51. doi: 10.23919/INDIACom49435.2020.9083712.
[40] S. Jagannathan, “FPGA Implementation of EPDT for Real-Time Removal of Salt and Pepper Noise,” 2011.
[41] P.-H. Lin, B.-H. Chen, F.-C. Cheng, and S.-C. Huang, “A Morphological Mean Filter for Impulse Noise Removal,” J. Disp. Technol., vol. 12, no. 4, pp. 344–350, 2016, [Online]. Available: https://opg.optica.org/jdt/abstract.cfm?URI=jdt-12-4-344.
[42] Setyobudi, R. (2023). Utilization of tds sensors for water quality monitoring and water filtering of carp pools using IoT. EUREKA: Physics and Engineering, (6), 69-77.
[43] Y. Jiao, C. Qian, and S. Fei, “Mask Convolution for Filtering on Irregular-Shaped Image,” in 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), 2018, pp. 115–118. doi: 10.1109/DCABES.2018.00039.