

DEEP LEARNING INDABA

A GAN Solution for Data imbalance in Atherosclerosis Screening Amel LAIDI^{1,*}, Mohammed AMMAR², Mostafa EL HABIB DAHO³

¹LIMOSE Laboratory M'Hamed Bougara University Boumerdes, Algeria
 ²LIST Laboratory, University M'Hamed Bougara, Boumerdes, Algeria
 ³Biomedical Engineering Laboratory, University of Tlemcen, Tlemcen, Algeria

Abstract

In our paper "GAN Data Augmentation for Improved Automated Atherosclerosis Screening from CCTA" [1], we introduced an end-to-end model for atherosclerosis detection straight from CCTA images. the model used transfer learning to overcome the issue of small training data, it achieved 95.2% accuracy, 60.8% sensitivity, and 90.48% PPV.

The model's performance was then improved by

Generative Adversarial Networks

Data Augmentation Using Generative Adversarial Networks (GAN) was first introduced in a paper on 2019 [6], The idea behind GAN is to generate images similar to the original data, enough to fool a discriminator.

training a GAN happens in two parts

• Part 1: The Discriminator is trained while the generator's parameters are fixed. the network only forward propagates. The Discriminator is

and a set of the set o	and the second se			and the second
the second se	and the second s	and the second se	State and Description of American Street, Stre	and the second
	THE OWNER WATER CONTRACTOR	And only the second states and the second states and the second states	THE OWNER AND ADDRESS OF TAXABLE PARTY.	A REAL PROPERTY OF A READ PROPERTY OF A REAL PROPER
- Statement water many or president the state of the second	TANK OF TAXABLE PARTY.	And with the second states and second s	THE OWNER AND ADDRESS OF TAXABLE PARTY.	- A REAL PROPERTY OF THE PARTY
- COMPANY AND ADDRESS OF THE OWNER	TAXABLE PROPERTY AND INCOME.	And and the second	TAXABLE PARTY OF TAXABLE PARTY.	- The second sec
	NAME OF TAXABLE PARTY.	And and Property of the party o	COLUMN TWO IS NOT THE OWNER. NAMED IN COLUMN	And in case of the local division of the loc
Contraction of the second s		A REAL PROPERTY AND A REAL	and the second division of the second divisio	and the second
The Constitution of the United States and the State	Statement of the statem	Provide and the August for the Automatical and Automatical Strategy and the	Statement of the Statement of the Statement	A second s
and the second se	CONTRACTOR OF THE OWNER WATER OF THE OWNER OWNER OF THE OWNER OWNE	CONTRACTOR OF THE OWNER	Statement Street Street Street Street	A DESCRIPTION OF THE OWNER OWN
- Strengt of the second se	CONTRACTOR OF THE OWNER.	And a second	Station over the state of the local division	CONTRACTOR OF THE OWNER
- State and South State and State and State and State and State	And the second s	ALC: AND REAL PROPERTY AND REAL PROPERTY AND REAL PROPERTY.	And in case of the local division of the loc	STATISTICS OF STREET, STRE
The statement of the second se	AND DECIMAL OF THE OWNER.	ALL CONTRACTOR OF A DESCRIPTION OF A DESCRIPANTE A DESCRIPANTE A DESCRIPANTE A DESCRIPTION OF A DESCRIPTIONO	COMPANY OF THE OWNER.	second of the local division of the local di
and the second	and the second division of the second divisio	Science and a second	Contraction in the owner of the owner owner of the owner	second seco
The second s	The second	and the second second second second for an of far stillering may if the better second second	and a state of the	The State of the S
The second s	Contraction of Contra	Construction of the second s	and the second se	the second se
A provide the statement of the statement os	And in case of the local division of the loc	And the second second second second second	And Personal Property in which the	and the second
A DESCRIPTION OF A DESC	AND DESCRIPTION OF THE OWNER.	And a lot of the second s	And Personal Property in which the real of the local division of t	succession of the local division of the loca
Automore in state from an operation in the sector of the s	And Description of the local division of the	envillement disease an annual second statement of the large	And Person Name and Post Office of the Owner, or other	And Address of Article Property and Address of The
And show the second sec	And Description of the local division of the	Provident of the Party of the P	And Personal Property Named and Post	and some of the subscription of the local division of the local di
The second s	NAME OF TAXABLE PARTY OF TAXABLE PARTY.	Contraction of the second difference were allowed with the second s	NAME AND ADDRESS OF TAXABLE PARTY.	and the second se
and the second	The second se	Service Chemica and December on California and December of Language Service California	the second se	the state of the second s

Figure 3. A sample of positive images from original dataset

Concernance of the local division of the loc	1 Personal Providence of the local division	COLUMN TWO IS NOT THE OWNER.	CONTRACTOR OF THE OWNER OWNER OF THE OWNER OWNE OWNER OWNE	THE OWNER WHEN THE OWNER
	A	No. of Concession, Name	A REAL PROPERTY AND INCOME.	THE R COMPANY OF
and the second s	1 CONTRACTOR		A CONTRACTOR OF TAXABLE PARTY.	1 THE R OWNER WHEN THE R.
and the second se	Commenter Bight	Concession of the local division of the loca	A CONTRACTOR OFFICE	A REAL PROPERTY AND ADDRESS OF TAXABLE PARTY.
the second se		Contraction of the local division of the loc	CONTRACTOR OFFICE	· strategies and a strategiest
	10000		· Contraction of Contract, or other	a sea second second line
	1 - BARRING BR	Contraction of the local division of the loc	· ····································	CONTRACTOR OF THE OWNER.
CONTRACTOR OF THE OWNER.	Contraction of the local division of the loc	Contraction and Contraction Advances	Contraction of the local division of the	Concentration of the local division of the l

Figure 4. A sample of positive images generated by the GAN

generating training data using a Generative Adversarial Network, to overcome the issue of class imbalance in the test set. The new model recorded a small drop in accuracy to 93.2%, but an improvement in overall performance with 89.0% sensitivity, and 97.13% PPV.

Keywords: Atherosclerosis, CCTA, Transfer learning, GAN, Data augmentation



Figure 1. Fat buildup in coronary arteries

- trained on a positive sample x from the real data set. while the generator generates a negative sample G(z).
- Part 2: The Generator is trained while discriminator's parameters are fixed. The generated data is fed into the discriminator, The error is calculated between the output of the discriminator D (G(z)) and the sample label, and the parameters of generator are updated using the error of backpropagation algorithm.
 The steps above are repeated until approaching Pdata(x) = Pg(x) . one the generator is well balanced

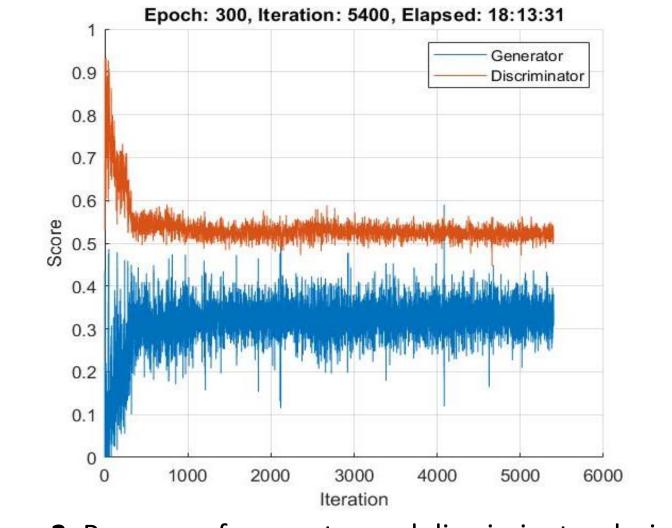
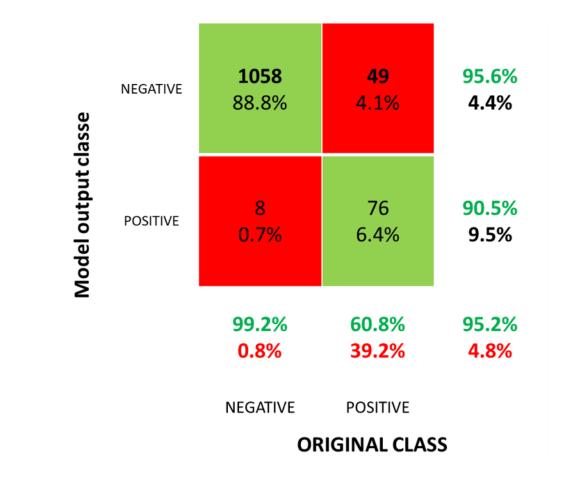


Figure 2. Progress of generator and discriminator during GAN

Discussion

- Intrinsic imbalance is a direct result of the nature of the dataspace, it is common in medical datasets as it reflects the prevalence of a disease in real life. It can affect the performance of deep learning algorithms.
- The performance of our model on the original dataset was comparable to Candemir et al. [8] who has achieved 90.9% accuracy, 68.9% Sensitivity, 93.6% specificity, and 58.8% PPV, and 96.1% NPV.



When new GAN generated were added to balance the dataset, the same model achieved

Introduction

Coronary artery disease generally, and atherosclerosis particularly, are the leading cause of world mortality [2].

Coronary artery disease is generally defined as the gradual narrowing of the lumen of the coronary arteries due to atherosclerosis; The buildup of plaque in the inner lining of the coronary arteries causes them to narrow in size and lose rigidity. CCTA is the preferred modality for evaluation of patients with coronary artery diseases. It is a non-invasive imaging technique that permits thorough description of coronary artery plaque and grading of coronary artery stenosis [3].

Class imbalance is a rather common issue in medical datasets; solutions include: simple image transformations, increasing the dataset size via data augmentation and image synthesis [4], random oversampling, random undersampling [5], and informed undersampling. training

GAN data augmentation for Imbalanced data

We used a pretrained Resnet on an open-source dataset [7] which achieved 95.2% accuracy, 60.8% sensitivity, 99.25% specificity and 90.48% PPV We noticed the test dataset has an uneven distribution of images (125 positive images vs. 1066 negative Images)

We then used GAN to generate new positive images and balance the dataset. We ran several tests to find the best number of images to add to the dataset, with the right balance of improved specificity without losing too much in accuracy.

The new approach achieved 93.2% of accuracy, 89.0% of sensitivity, 97.37% of specificity ,97.13% of PPV and 89.9% of NPV.

Table 1. the outcome of testing with generated images

93.2% of accuracy, 89.0% of sensitivity, 97.37% of specificity ,97.13% of PPV and 89.9% of NPV. The model got better at recognizing positive

89.9% 1038 117 NEGATIVE 10.1% 48.7% 5.5% 949 97.1% POSITIVE 44.5% 2.9% 1.3% 93.2% 6.8% 11.0% **ORIGINAL CLASS**

cases.

Conclusions

GAN is a useful tool for data augmentation, and it worked well on our datatset to overcome the imbalanced data issue.

The results achieved are not only comparable to the state of the art, but also allow a fast and accurate automatic detection of atherosclerosis, which has

Added images	Accuracy	sensitivity	specifity	PPV	NPV
Original	95.2	60.8	99.25	90.48	95.57
100 images	94.0	75.2	96.2	60.6	97.1
300 images	86.9	81.6	87.5	43.4	97.6
1000 images	93.2	89.0	97.4	97.1	89.9

important clinical implications.

For a more detailed description of the method and results, check the paper (or scan QR code below)

Contact

AMEL LAIDI

LIMOSE Laboratory, MHAMED BOUGARA UNIVERISTY

Email: a.laidi@univ-Boumerdes.dz

Linkedin: www.linkedin.com/in/amel-laidi-

Phone: +213662817554

References

- A. Laidi, M. Ammar, M. E. H. Daho and S. Mahmoudi, "GAN Data Augmentation for Improved Automated Atherosclerosis Screening from Coronary CT Angiography," EAI Endorsed Transactions on Scalable Information Systems: Online First, 2022.
- "World Health Organization," 11 June 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).
 [Accessed 23 May 2022].
- O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester and others, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?," IEEE transactions on medical imaging, vol. 37, no. 11, pp. 2514--2525, 2018.
- 4. L. Engstrom, B. Tran, D. Tsipras, L. Schmidt and A. Madry, "A rotation and a translation suffice: Fooling cnns with simple transformations," arXiv, 2018.
- A. Fernàndez, S. García, M. Galar, R. C. Prati, B. Krawczyk and F. Herrera, Learning from imbalanced data sets, vol. 10, Springer, 2018.
- F. H. K. d. S. Tanaka and C. Aranha, "Data augmentation using GANs," arXiv preprint arXiv:1904.09135, 2019.
- M. Demirer, V. Gupta, M. Bigelow, B. Erdal, L. Prevedello and R. White, "Image dataset for a CNN algorithm development to detect coronary atherosclerosis in coronary CT angiography," 2019.
- 8. S. Candemir, R. D. White, M. Demirer, V. Gupta, M. T. Bigelow, L. M. Prevedello and 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, vol. 83, p. 101721, 2020.

