POLYTECH MONTPELLIER

Data science,
Management &
Software architecture
Promotion 2025–2026

LABORATOIRE
de MÉCANIQUE
et GÉNIE CIVIL

3D Microstructural Analysis of Arteries Using Artificial Intelligence

Final-Year Engineering Project Report

University of Montpellier
Polytech Montpellier
Department of Data science, Management & Software architecture (DaMS5)

Students:

Maxime Dudognon
Matthéo Dascalu
Hugo Brun

Scientific Supervisors:

Cristina Cavinato
LMGC (Laboratoire de Mécanique
et Génie Civil)

Academic Tutor:

Christophe Fiorio (Polytech)

Gérard Subsol
LIRMM (Laboratoire d'Informatique,
de Robotique et de Microélectronique)

December 2025 – February 2026

Version of February 3, 2026

Acknowledgments

We would like to express our sincere gratitude to our scientific supervisors, Cristina Cavinato and Gérard Subsol, for their invaluable guidance, constructive feedback, and continuous support throughout this project. We also thank our academic tutor, Christophe Fiorio, for his dedicated mentorship and for facilitating communication with the ISDM-MESO computing platform, which was essential in obtaining access to high-performance computing resources.

We acknowledge the support of the ISDM-MESO HPC platform and are particularly grateful to Luca Cimini, system and network administrator at Polytech Montpellier, for his assistance with the computational infrastructure at Polytech Montpellier, as well as to Anne Laurent, director of the ISDM, for her significant help in securing and managing access to ISDM resources.

Contents

Figures

1 Introduction

Cardiovascular diseases are the leading cause of death worldwide. The World Health Organization reports about 19.8 million deaths in 2022, which is around 32 % of all global deaths [7]. Within this broad category, acute aortic diseases (such as aortic dissection) are especially dangerous. A summary of results from the International Registry of Acute Aortic dissection (IRAD) reports an in-hospital mortality of about 22 % for type A dissections, and up to 57 % for patients treated with medical therapy only [4].

To better understand why these events happen, researchers study the structure of the aortic wall. From a mechanical point of view, the aorta works like an elastic reservoir: it expands during systole, stores elastic energy, and then recoils during diastole to help maintain blood flow [2]. This behavior depends on the aorta's microstructure, mainly elastic fibers (elastin-based structures) and collagen fibers, which are the main load-bearing components of the tissue [1].

A key difficulty is that microstructural damage can persist for a long time. Elastic fibers renew very slowly (reported on the order of decades, and elastin has been described as having a biological half-life of about 70 years), so degradation tends to accumulate and can have long-term mechanical effects [1, 6]. In the Marfan aorta, experimental results show that when elastic lamellae become more porous, the aorta stores less elastic energy and becomes stiffer [1]. In addition, biological signaling pathways can actively contribute to disease progression; for example, fibronectin–integrin $\alpha 5$ signaling has been identified as an important driver of thoracic aortic aortic aneurysm in a Marfan mouse model [3].

Modern 3D optical imaging methods provide detailed views of arterial microstructure. For example, multiphoton-microscopy can acquire 3D image stacks with sub-micron in-plane sampling and micron-scale depth steps, giving rich volumetric data about collagen, elastin, and cellular organization [2]. However, these datasets are large and complex, which makes manual analysis slow and difficult to scale.

This motivates machine learning approaches that can learn relationships between microstructure and mechanical behavior. Prior work has shown that deep learning can predict arterial mechanical behavior from microstructural data with strong agreement to experiments (for example, reported $R^2 = 0.92$ and a maximum error below 10 %) [6].

This project was conducted in collaboration with the Laboratoire de Mécanique et Génie Civil (LMGC) and the Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM). The research aims to develop computational tools to support ongoing investigations into arterial biomechanics and disease mechanisms, with the ultimate goal of improving risk prediction for aortic pathologies. In this context, the goal of this final-year project is to evaluate whether machine learning methods are feasible on an internal dataset of 3D multichannel microscopy volumes from mouse arterial tissue, given practical constraints such as limited annotations and high data dimensionality. Specifically, this project delivers: (1) a complete data processing pipeline to handle 650 GB of raw multiphoton-microscopy microscopy data and construct 772 standardized volumetric stacks; (2) implementation and evaluation of 3D convolutional neural network architectures for binary disease classification; (3) a critical analysis of the feasibility, limitations, and challenges of applying deep learning to small-sample biomedical imaging datasets.

The remainder of this report is organized as follows. Section 2 describes the dataset characteristics, preprocessing pipeline, and data quality assessment. Section 3 presents the deep learning architectures explored, training strategies, and experimental setup. Section 4 details the team structure, project timeline, development infrastructure, and methodology. Section 5 provides a critical analysis of technical and organizational choices, challenges overcome, and environmental considerations. Finally, Section 6 summarizes contributions, personal learnings, and future perspectives.

2 Data

2.1 Raw dataset overview

The dataset used in this project consists of three-dimensional microscopy data acquired from **mouse arterial tissue**. The raw data were initially provided on a single external hard drive containing approximately **650 GB** of unprocessed files. The data layout was non-standard and consisted of deeply nested directory structures, where a significant portion of the experimental metadata was encoded directly in folder names. In addition, complementary metadata were provided in the form of spreadsheet files, which had to be programmatically linked to the corresponding image data.

Each arterial sample corresponds to a volumetric acquisition split into individual two-dimensional image files. In total, the dataset contained approximately **250,000 TIFF files** (image format), stored on a slow storage medium, which made any direct computation or inspection impractical prior to migration to a HPC (high-performance computing) environment.

Each image file corresponds to one slice of a three-channel acquisition, with channels encoded using RGB conventions:

- **Red channel**: collagen fibers,
- **Green channel**: elastin,
- **Blue channel**: cell nuclei.

2.2 Metadata and sample information

Each volumetric sample is associated with a set of biological and experimental metadata, including genetic background, anatomical region, mechanical loading conditions, and disease status. Figure 1 presents the metadata fields and their meanings, with an example from a representative sample.

Field	**Description**	**Example**
Classe	Disease status (healthy/diseased)	MALADE
Genetic	Genetic modification or strain variant	alpha5/alpha2, mgR
Sex	Biological sex of the specimen	M
Strain	Mouse strain	C57
Region	Anatomical region of the artery	ATA
Age (wk)	Age of the specimen in weeks	9
Pressure	Applied pressure in mmHg	120
Axial Stretch	Axial stretch ratio	2
Orientation	Sample orientation (systole/diastole)	D

Figure 1: Metadata structure and example values for volumetric samples

These metadata were not used as primary learning targets, but were instead leveraged to ensure balanced data splits and to support exploratory analyses. A small number of samples

were excluded after consultation with the domain expert due to missing labels or poor image quality.

2.3 Data access and computing infrastructure

Due to the large number of small files and the very limited I/O performance of the original storage medium, initial data exploration was effectively infeasible. Access to appropriate computing resources was therefore a prerequisite for any meaningful processing.

All large-scale preprocessing and dataset construction were performed on the **IO cluster**, part of the ISDM-MESO HPC (high-performance computing) platform [5]. Gaining access to this infrastructure required approximately one month, during which no significant computation could be performed. Additional experiments were later conducted on internal Polytech computing resources, consisting of an internal teaching cluster with exclusive access to 32 CPU cores and a single NVIDIA L40S GPU.

2.4 Data views and reproducibility strategy

A strict design choice was made to **never modify the original raw data directly**. Instead, all preprocessing steps were performed on derived views of the dataset, implemented using symbolic links to the original files. This approach ensured full traceability, reproducibility, and consistency across collaborators, while allowing multiple iterations of preprocessing without risking data corruption.

Multiple successive views were created to progressively standardize folder names, extract metadata automatically from directory hierarchies, and remove irrelevant or malformed samples. This strategy also facilitated direct communication with the biological expert, as references to the original data structure remained valid throughout the project.

2.5 Stack construction

After data validation, volumetric stacks were reconstructed from individual TIFF files and stored in NIfTI (.nii.gz) format. Each stack follows a consistent axis convention `(Z, Y, X, C)`, where $C = 3$ corresponds to the 3 channels (see Figure 2).

Channel	Protein/Structure	Contrast Window
Red	Collagen	300–1,800
Green	Elastin	170–300
Blue	Cell nuclei	500–4,000

Figure 2: Multiphoton microscopy channel composition and visualization contrast settings applied in ITK-SNAP

A total of **772 volumetric stacks** were successfully generated. Each stack has a spatial resolution of approximately $1042 \times 1042 \times D \times 3$, with D varying between samples. The dataset construction process was fully logged using structured JSON metadata files, ensuring traceability from raw files to final volumes.

2.6 Data distribution and intensity characteristics

All reconstructed volumes exhibit a similar global structure, although the image quality varies between samples. The raw intensity values range from approximately 0 to 16,000, with the majority of values concentrated between 0 and 2,000. Figure 3 shows a representative slice visualized in ITK-SNAP, displaying all three channels simultaneously. To optimize visualization and highlight the relevant biological structures, the contrast window for each channel was adjusted as detailed in Figure 2.

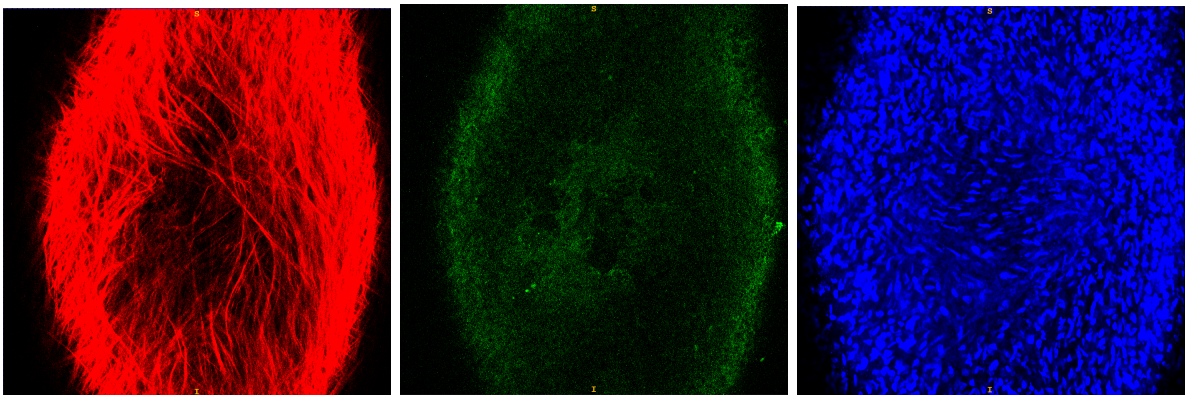

Figure 3: Collagen, elastin and cell nuclei slices of an example volume

2.7 Patch extraction and preprocessing

Due to the large spatial size of the volumes, a patch-based strategy was adopted. For each volume, an average pooling operation was applied using a kernel of size $(256, 256, 32)$ to identify regions containing the highest signal intensity. Around the maxima of this pooled response, a fixed set of eight patches of shape $(256, 256, 32, 3)$ was deterministically extracted per volume.

All patches were normalized using a global z-score normalization computed across *all patches and all volumes*.

It is important to emphasize that this patch-based strategy does **not** increase the number of independent labels available for learning. Although each volume yields multiple patches, all patches inherit the same global label. As a result, the effective number of labeled samples remains equal to the number of volumes, not the number of patches. Patch extraction improves computational tractability and spatial focus, but does not address label scarcity or statistical limitations.

2.8 Labels and learning targets

The only label available for all samples is a binary classification indicating disease status (`SAIN` / `MALADE`). The dataset is approximately balanced with respect to this label. Additional quantitative labels are available for a very small subset of samples (20), including measurements such as volume fractions, fiber ondulation, and cell density. Figure 4 presents the structure of these quantitative measurements with example values from a representative sample.

Measurement	Description	Example
In Vivo Stretch		
Axial	Longitudinal stretch ratio	1.38
Circumferential	Radial stretch ratio	1.71
Unloaded Thickness		
Thickness	Wall thickness in micrometers	73 µm
Volume Fraction		
Collagen	Percentage of collagen in tissue	28.56%
Elastin	Percentage of elastin in tissue	28.00%
Cell	Percentage of cellular content	43.44%
Cell Density		
Adventitia	Cell count per 0.001 mm^3 in outer layer	30.73
Media	Cell count per 0.001 mm^3 in middle layer	382.91
Intima	Cell count per 0.01 mm^2 in inner layer	6.67
Fiber Ondulation		
Straightness	Fiber straightness index (0-1)	0.96
Fiber Bundle Width		
Width	Average bundle width in micrometers	16.12 µm

Figure 4: Quantitative measurement structure and example values (available for 20 samples)

These labels cover only a limited portion of the dataset and do not span the full range of biological variability. Consequently, they cannot support statistically robust supervised learning or reliable validation. In this project, these additional labels are therefore considered **exploratory** and are used only for qualitative analysis or proof-of-concept experiments.

2.8.1 Label distribution statistics and correlation analysis

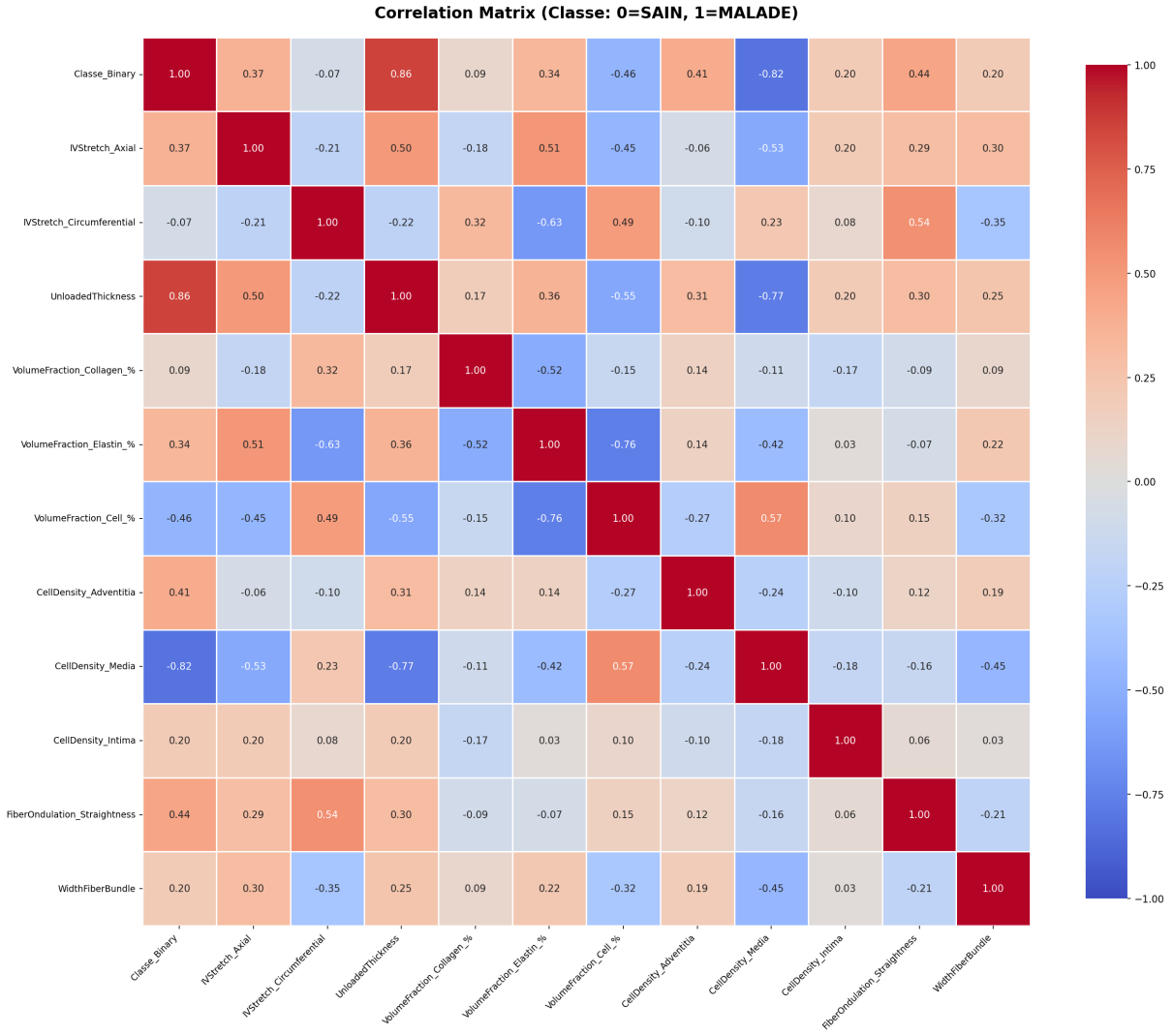

Figure 5: Correlation Matrix of quantitative labels and disease status ($N = 20$). Pearson correlation coefficients are shown, with `Classe_Binary` representing disease status ($0 = \text{HEALTHY}, 1 = \text{DISEASED}$).

The correlation matrix in Figure 5 reveals significant biological relationships:

- **Disease Indicators:** `UnloadedThickness` shows a very strong positive correlation with disease status ($r = 0.86$), while `CellDensity_Media` shows a strong negative correlation ($r = -0.82$). This indicates that disease is likely associated with thicker tissue and reduced medial cell density.
- **Structural Relationships:** Strong internal correlations exist between metrics, such as the negative correlation between `UnloadedThickness` and `CellDensity_Media` ($r = -0.77$). This suggests that increased wall thickness is associated with a loss or redistribution of medial cells, consistent with structural remodeling.

2.8.2 Exploratory visualization of secondary quantitative labels

Due to the limited sample size ($N = 20$), these quantitative measurements serve to confirm that the captured image features align with expected pathological remodeling.

3 Deep Learning Approaches

4 Project Organization

4.1 Team structure and roles

The project was conducted by a team of three final-year engineering students from Polytech Montpellier: Maxime Dudognon, Matthéo Dascalu, and Hugo Brun. The work was carried out primarily in person at Polytech facilities, fostering close collaboration and rapid problem-solving.

4.1.1 Role distribution

While the team operated collaboratively with all members contributing to all aspects of the project, certain areas of focus naturally emerged based on interests and prior experience:

- **Maxime Dudognon**: Coordinated overall project progress and contributed technical expertise from prior medical imaging experience gained during an internship at EPFL.
- **Matthéo Dascalu**: Primary focus on data analysis, exploratory visualization, and interpretation of results in the biological context.
- **Hugo Brun**: Led data preprocessing pipeline development, volumetric stack construction, and model training experiments.

Maxime brought relevant practical experience from a previous internship at EPFL (École Polytechnique Fédérale de Lausanne) in Switzerland, where he worked on medical deep learning imaging. Matthéo and Hugo had foundational knowledge from academic coursework and personal projects but gained most of their practical deep learning experience through this project.

The team adopted a flexible approach where technical decisions were made collectively, and tasks were distributed based on availability and individual interest rather than rigid assignments. This collaborative model ensured knowledge sharing and allowed each member to develop competencies across the full project scope.

4.1.2 Supervision and guidance

The project operated under dual scientific supervision:

- **Cristina Cavinato (LMGC)**: Provided domain expertise in arterial biomechanics, validated biological interpretations of results, and guided decisions related to data quality and experimental metadata.
- **Gérard Subsol (LIRMM)**: Offered technical guidance on machine learning methodologies, model architecture selection, and computational strategies.
- **Christophe Fiorio (Polytech Montpellier)**: Served as academic tutor, ensuring alignment with pedagogical objectives and institutional requirements.

The supervision structure evolved over the course of the project. Initial meetings involved all supervisors to establish a common understanding of objectives and scope. The first meeting included Cristina Cavinato, Gérard Subsol, and Christophe Fiorio; the second meeting included

Cristina Cavinato and Christophe Fiorio. Following these initial sessions, communication transitioned to a more asynchronous model, with email updates to keep supervisors informed of progress and requests for guidance when technical or biological questions arose. Meetings with individual supervisors were arranged as needed, typically when major decisions required validation or when unexpected challenges emerged.

The team operated with significant autonomy. While supervisors were consulted before initiating major new directions (e.g., switching from one type of prediction task to another), day-to-day technical decisions and implementation choices were made independently by the team.

4.1.3 Collaboration infrastructure

To facilitate coordination, the team utilized shared computing resources:

- A common virtual machine on the Polytech internal cluster, providing centralized access to data, code, and experimental results.
- Shared workspace on the ISDM-MESO HPC cluster for large-scale preprocessing and training.

Clear communication and regular in-person synchronization prevented conflicts and ensured consistency across parallel workstreams. This setup allowed the team to work efficiently while maintaining full traceability and reproducibility of all experimental steps.

4.2 Timeline and milestones

Weeks	Phase & Activities	Key Event
1-2	Initial setup: meetings, literature review	Dec 15: Data received
2-4	Data preprocessing: validation, symbolic links, metadata	
4-5	Infrastructure: HPC access, stack construction (772 volumes)	Jan 9-12: HPC access
5-8	Model development: architecture, training, tuning	
8-10	Analysis & documentation: results, report, presentation	Feb 18: Final presentation

Figure 6: Project timeline and major phases

The project spanned approximately **10 weeks** (December 8, 2025 - February 18, 2026) at 35 hours per week per team member. Figure 6 illustrates the main phases and their dependencies.

4.2.1 Phase 1: Initial setup (Weeks 1-2)

Two in-person meetings established the scope and objectives of the project. The review of the literature began immediately. The external hard drive arrived on December 15, one week after the start of the project.

4.2.2 Phase 2: Data preprocessing (Weeks 2-4)

Initial data exploration revealed organizational challenges. Preprocessing scripts were developed to validate, clean, and structure the dataset using symbolic link views. Metadata extraction from folder names and spreadsheets was automated. Work proceeded on local machines with limited computational capacity.

4.2.3 Phase 3: Infrastructure and stack construction (Weeks 4-5)

Access to HPC resources was granted on January 9 (ISDM-MESO: CPU and storage) and January 12 (Polytech cluster: virtual machine with 32 CPU cores, NVIDIA L40S GPU and storage). Dataset migration enabled full-scale stack construction, yielding 772 NIfTI (.nii.gz) volumes. A mid-project video call with Cristina Cavinato confirmed alignment with biological objectives.

4.2.4 Phase 4: Model development (Weeks 5-8)

With the infrastructure operational, 3D convolutional architectures were implemented and trained. ???

4.2.5 Phase 5: Analysis and documentation (Weeks 8-10)

???

4.3 Development tools and infrastructure

4.3.1 Computing resources

The project relied on two complementary HPC platforms:

ISDM-MESO cluster Accessed via SSH and SLURM scheduler, this platform provided CPU resources and distributed storage. It was primarily used for initial data preprocessing, stack construction, and metadata extraction, and continued to serve as auxiliary storage throughout the project.

Polytech internal cluster A dedicated virtual machine with 32 CPU cores, 48 GB RAM, one NVIDIA L40S GPU, and approximately 3 TB of storage. Accessed via SSH, this became the primary environment for all GPU-based model training and served as the main development platform.

Preprocessing tasks were distributed across both platforms, while all deep learning training occurred exclusively on the Polytech cluster due to GPU requirements.

4.3.2 Software stack

The project was implemented in **Python 3.12** using **PyTorch** as the deep learning framework. Key libraries included:

- **Data processing**: NumPy, nibabel (NIfTI (.nii.gz) files), Pandas (metadata handling)

- **Visualization**: ITK-SNAP (3D volume inspection), Matplotlib (plots and metrics), Napari and ImageJ (supplementary visualization)
- **Experiment tracking**: Weights & Biases (W&B) for automated logging, supplemented by manual logs
- **Environment management**: Shared Conda environment synchronized across both platforms

4.3.3 Collaboration infrastructure

Code and data were shared through directories on the Polytech VM and ISDM-MESO workspace. The codebase combined Python scripts for preprocessing and training pipelines with Jupyter notebooks for exploratory analysis.

Team communication occurred primarily in person, with an Instagram group and Discord calls for remote coordination. Supervisor communication used email and Zoom. This report was collaboratively written in Overleaf.

4.4 Workflow and methodology

The team adopted a pragmatic and iterative development approach, balancing careful planning with rapid prototyping. Daily in-person collaboration at Polytech facilities enabled continuous informal communication, while formal meetings were organized as needed to discuss strategic decisions or technical roadblocks.

4.4.1 Development cycle

For complex tasks, the team prioritized planning before implementation. Preprocessing pipelines and architectural choices were designed based on literature review and feasibility analysis before coding began. However, when uncertainty was high, quick prototyping on small data subsets was used to validate assumptions before committing to full-scale implementation. This hybrid approach-structured planning combined with exploratory testing-proved effective for navigating the project's technical and biological constraints.

Code review was systematically performed within the team to ensure consistency and catch errors early. All members had access to shared code repositories on the Polytech VM, facilitating peer review and collaborative debugging.

4.4.2 Validation and quality assurance

Preprocessing correctness was verified through a combination of manual inspection and quantitative metrics. Visual checks using ITK-SNAP ensured that volumetric stacks were correctly reconstructed and that intensity normalization preserved biological structures. When biological plausibility was uncertain, the team consulted Cristina Cavinato for validation.

Model performance was assessed using F1 scores across different preprocessing configurations, allowing the team to empirically identify the most effective data preparation strategies. This

metric-driven approach ensured that preprocessing choices were grounded in measurable outcomes rather than intuition alone.

4.4.3 Experiment tracking and iteration

All training runs were systematically logged using Weights & Biases (W&B). Both successful and failed experiments were documented to avoid redundant work and to build institutional knowledge within the team. This comprehensive logging proved essential when diagnosing issues or comparing alternative approaches.

Architecture selection followed a structured methodology: initial experiments on reduced datasets tested a range of models (ResNet3D variants, EfficientNet3D, ViT3D, ConvNeXt3D, SE-ResNet3D), informed by literature review. Results from these small-scale tests guided decisions on which architectures warranted full-scale training, balancing computational cost with expected performance gains.

4.5 Challenges and solutions

4.5.1 Infrastructure access delays

The one-month delay in obtaining access to HPC resources posed an initial obstacle to large-scale computation. Rather than waiting passively, the team used this period productively for literature review, identifying relevant architectures and preprocessing strategies from prior work on 3D medical imaging. Preprocessing scripts were developed and tested on local machines using small data samples, ensuring that when infrastructure became available, the pipeline was ready for immediate deployment. This proactive approach minimized the impact of the delay on overall project progress.

4.5.2 Non-standard data organization

The raw dataset's deeply nested directory structure and metadata encoded in folder names presented a data engineering challenge. Cristina Cavinato clarified the organizational logic during the second meeting, which accelerated understanding. The team's solution-creating symbolic links to the original files rather than duplicating data-was conceptually straightforward but required careful implementation to ensure reproducibility and traceability. This approach proved highly effective, preserving data integrity while enabling flexible reorganization for processing needs.

4.5.3 Label scarcity and learning strategy

The limited availability of quantitative annotations (only 20 samples with detailed measurements) constrained the project's scope. The team explored self-supervised learning as a potential workaround but ultimately focused efforts on binary disease classification, where labels were available for all samples. The quantitative labels were retained for exploratory analysis but were not integrated into the primary learning pipeline.

5 Analysis and Comments

5.1 Technical choices and justification

5.2 Organizational effectiveness

5.3 Challenges and their resolution

5.4 Environmental and social responsibility

6 Conclusion

6.1 Contributions and deliverables

6.2 Personal and professional development

6.3 Limitations and future perspectives

Glossary

Aortic aneurysm An abnormal enlargement or bulging of the aortic wall, often resulting from structural weakening due to disease or genetic factors. 5, 22

Aortic dissection A serious medical condition where a tear occurs in the inner layer of the aorta, allowing blood to flow between the layers of the aortic wall. 5, 22

Average pooling A downsampling operation that reduces spatial dimensions by computing the mean value within a sliding kernel, used to identify high-intensity regions. 9

Axial stretch The ratio of stretched length to original length along the longitudinal axis of a vessel or tissue sample. 7

Cell nuclei The central organelles of cells containing genetic material (DNA); visualized in the blue channel of the multiphoton microscopy data to identify cellular distribution within arterial tissue. 7, 8

Collagen A structural protein that provides tensile strength to arterial tissue; visualized in the red channel of the microscopy data. 5, 7, 8, 10

CPU Central Processing Unit; the main processor of a computer that executes instructions. 8, 16

Diastole The relaxation phase of the cardiac cycle, during which the heart fills with blood. 5, 7

Elastin An elastic protein that allows arterial tissue to stretch and recoil; visualized in the green channel of the microscopy data. 5, 7, 8, 10, 22

GPU Graphics Processing Unit; a specialized processor originally designed for graphics rendering, now widely used for parallel computation in machine learning. 8, 16

HPC High-Performance Computing; specialized computing infrastructure designed for processing large-scale computations and data-intensive tasks. 7, 8, 15, 16, 18

I/O Input/Output; refers to data transfer operations between a computer and external storage devices. 8

IRAD International Registry of Acute Aortic Dissection; a multicenter database tracking clinical outcomes of acute aortic dissection cases worldwide. 5

ISDM-MESO Institut des Sciences des Données de Montpellier - MESO computing platform; a high-performance computing infrastructure providing computational resources and data storage for research projects in the Montpellier area. 8, 15–17

ITK-SNAP An open-source software application for 3D medical image segmentation and visualization, providing interactive tools for viewing and annotating volumetric data. 9, 17

JSON JavaScript Object Notation; a lightweight data format for storing and exchanging structured information, commonly used for metadata storage. 8

Kernel A small matrix used in image processing operations such as convolution or pooling to extract features or reduce dimensionality. 9, 21

Marfan Refers to Marfan syndrome, a genetic disorder affecting connective tissue, particularly elastin and fibrillin, leading to increased risk of aortic aneurysm and aortic dissection. Used as an adjective to describe affected tissues or animal models. 5

Metadata Data that provides information about other data; in this project, includes experimental conditions, genetic information, and anatomical details. 7, 8, 14, 16

Multiphoton microscopy An advanced optical imaging technique that uses nonlinear excitation to capture high-resolution 3D images of tissue microstructure. 5

NIfTI (.nii.gz) Neuroimaging Informatics Technology Initiative format; a file format commonly used for storing volumetric stack medical imaging data, with .gz indicating gzip compression. 8, 16

Patch A small 3D sub-volume extracted from a larger volumetric stack image, used to make computation tractable. 9

R^2 Coefficient of determination; a statistical measure indicating the proportion of variance in the dependent variable predictable from the independent variable(s), ranging from 0 to 1. 5

Symbolic link A file system reference that points to another file or directory, allowing access to data without duplicating it. 8, 18

Systole The contraction phase of the cardiac cycle, during which the heart pumps blood into the arterial system. 5, 7

TIFF Tagged Image File Format; a flexible file format for storing raster graphics and images, commonly used in microscopy. 7, 8

Volumetric stack A 3D image volume composed of multiple 2D image slices stacked along the depth axis. 5, 8, 17, 22

Weights & Biases (W&B) A machine learning platform for experiment tracking, visualization, and model management, commonly used to log metrics, hyperparameters, and artifacts during training. 17, 18

Z-score normalization A normalization method that transforms data to have zero mean and unit variance, calculated as $(x - \mu)/\sigma$. 9

References

[1] Cristina Cavinato, Hengdi Chen, Daniel Weiss, Juan-Carlos Ruiz-Rodríguez, Sherif Schwartz, and Jay D. Humphrey. Progressive microstructural deterioration dictates evolving biomechanical dysfunction in the Marfan aorta. *Frontiers in Cardiovascular Medicine*, 8:800730, 2021.

[2] Cristina Cavinato, Syed Imran Murtada, Andrés Rojas, and Jay D. Humphrey. Evolving structure-function relations during aortic maturation and aging revealed by multiphoton microscopy. *Mechanisms of Ageing and Development*, 196:111471, 2021.

[3] Rui Chen, Katharina Speichinger, Iolanda Ratera, Jin Chou, Christian Reeps, Julie Descamps, Julia Fuchs, Matthias Kretz, Miriam Baumann, Leire Hernandez, Judith Stenger, Manuel Biermann, Ayan Dasgupta, Christian Berwanger, and Lars Maegdefessel. Fibronectin (fn)–integrin $\alpha 5$ signaling promotes thoracic aortic aneurysm in a mouse model of marfan syndrome. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 43:e132–e150, 2023.

[4] Arturo Evangelista, Eric M. Isselbacher, Eduardo Bossone, Thomas G. Gleason, Marco Di Eusanio, Udo Sechtem, Michael P. Ehrlich, Stefano Trimarchi, Alan C. Braverman, Tor Myrmel, Kent M. Harris, Stuart Hutchison, Christoph A. Nienaber, David G. Montgomery, James B. Froehlich, Kim A. Eagle, and Lars A. Pape. Insights from the international registry of acute aortic dissection: a 20-year experience of collaborative clinical research. *Global Cardiology Science and Practice*, (3):8, 2016.

[5] ISDM-MESO HPC Platform. Isdm-meso high performance computing platform. Internal HPC infrastructure, 2025. Computations were performed on the ISDM-MESO HPC platform (IO cluster), funded in the framework of State–Region Planning Contracts (Contrat de plan État–Région, CPER) by the French Government, the Occitanie / Pyrénées-Méditerranée Region, Montpellier Méditerranée Métropole, and the University of Montpellier.

[6] Kevin Linka, Cristina Cavinato, Jay D. Humphrey, and Christian J. Cyron. Unraveling microstructure mechanics by bidirectional deep learning architecture. *Acta Biomaterialia*, 147:63–72, 2022.

[7] World Health Organization. Cardiovascular diseases (cvds), 2025. Accessed: 2026-01-22.

Résumé

Les maladies cardiovasculaires demeurent la première cause de mortalité mondiale, les pathologies aortiques aiguës telles que la dissection aortique présentant des taux de mortalité particulièrement élevés. La compréhension de l'organisation microstructurale du tissu artériel est essentielle pour prédire les défaillances mécaniques et la progression des maladies. Ce projet explore la faisabilité d'appliquer des méthodes d'apprentissage profond à des données de microscopie multiphotonique 3D de tissus artériels murins afin de prédire l'état pathologique à partir de caractéristiques microstructurales.

Le jeu de données comprend 772 volumes acquis à partir d'échantillons artériels, chacun contenant trois canaux visualisant le collagène, l'élastine et les noyaux cellulaires. En raison de contraintes computationnelles, une stratégie basée sur l'extraction de patches a été adoptée. Le projet a relevé des défis importants incluant une organisation non-standard des données, un accès limité aux infrastructures et une rareté des annotations.

Ce travail démontre la faisabilité technique du traitement et de l'analyse de données d'imagerie biomédicale 3D à grande échelle par apprentissage automatique, tout en soulignant l'importance critique de la qualité des données, de la planification infrastructurelle et de l'expertise métier dans les projets d'imagerie médicale.

Mots-clés

Apprentissage profond, Imagerie médicale, Microstructure artérielle,
Réseaux convolutifs 3D, Microscopie multiphotonique,
Dissection aortique, Calcul haute performance

Summary

Cardiovascular diseases remain the leading cause of death worldwide, with acute aortic diseases such as aortic dissection presenting particularly high mortality rates. Understanding the microstructural organization of arterial tissue is critical to predicting mechanical failure and disease progression. This project explores the feasibility of applying deep learning methods to 3D multiphoton microscopy data of mouse arterial tissue to predict disease status from microstructural features.

The dataset consists of 772 volumetric stacks acquired from arterial samples, each containing three channels visualizing collagen, elastin, and cell nuclei. Due to computational constraints, a patch-based strategy was adopted. The project addressed significant challenges including non-standard data organization, limited infrastructure access, and label scarcity.

This work demonstrates the technical feasibility of processing and analyzing large-scale 3D biomedical imaging data using machine learning approaches, while highlighting the critical importance of data quality, infrastructure planning, and domain expertise in medical imaging projects.

Keywords

Deep Learning, Medical Imaging, Arterial Microstructure,
3D Convolutional Networks, Multiphoton Microscopy,
Aortic Dissection, High-Performance Computing