MSc Biomedical Engineering (BME)
Final Project

Novel Deep Learning Methods for Modeling and Prediction of Abdominal Aortic Aneurysm Growth

Marieke Hofman

Supervisors:

Chair:	Prof. Dr. Christoph Brune
Daily:	Dieuwertje Alblas, MSc
	Dr. Jelmer M. Wolterink
External:	Dr. Erik Groot Jebbink

July, 2023

Mathematics of Imaging and AI (MIA)
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente

UNIVERSITY OF TWENTE. | TECHMED CENTRE | Mathematics of Imaging & AI

ABSTRACT

Abstract. Abdominal aortic aneurysms (AAAs) are focal dilatations of the abdominal aorta that, if left untreated, can rupture with high mortality rate. In this work, we present novel deep learning methods for modeling and prediction of local AAA growth. Using implicit neural representations (INRs), we obtained personalized, continuous representations of AAA shapes evolving over time, based on highly sparse and irregularly spaced longitudinal image data. We represent the AAA's outer wall evolving over time as the zero level set of its signed distance function (SDF), which we embed in a multilayer perceptron (MLP) that operates on space and time. We optimize this INR using automatically extracted AAA segmentations in longitudinal CTA data. This network is conditioned on spatio-temporal coordinates and therefore represents the evolving AAA shape at any spatial resolution and any point in time. Using regularization on spatial and temporal gradients of the SDF, we observe that our model can accurately interpolate AAA shapes evolving over time, with average surface distances (ASDs) ranging from 0.627 to 4.443 mm. This personalized approach for modeling AAA evolution, however, does not generalize easily to new AAA patients, limiting its adoption in clinical practice. To address this, we propose a graph convolutional network (GCN) for prediction of local AAA growth, operating on surface mesh representations of the AAA's outer wall. We optimize this network using continuous representations of evolving AAA shapes from multiple patients, that we obtained using the INRs. By conditioning the GCN model on a time step, we can predict AAA growth over any desired future time point. We demonstrate the GCN's performance to predict AAA shapes with diameter profiles along the AAA centerlines. The results indicate that our model can predict local AAA growth in the right direction specifically in the dilated part of the aorta, leaving the healthy parts unaffected. Our proposed pipeline, including automatic segmentation, continuous AAA surface representation, and predicting local AAA growth on the surface, holds potential clinical value for more personalized, pro-active AAA surveillance.

Keywords: Abdominal aortic aneurysm · Implicit neural representation · Graph convolutional network · Deep learning · Aneurysm growth

CONTENTS

Abstract **2**

1 Introduction **5**

2 Modeling AAA Evolution using Implicit Neural Representations **10**
2.1 Shape Representations . . . 10
2.2 Time-dependent Signed Distance Function . . . 11
2.3 Implicit Neural Representations . . . 11
2.3.1 Optimizing an INR . . . 11
2.4 Data . . . 13
2.4.1 Synthetic Dataset . . . 13
2.4.2 Aneurysm Dataset . . . 14
2.5 Experiments and Results . . . 16
2.5.1 Continuous Sphere Representation . . . 17
2.5.2 Continuous AAA Representation . . . 20
2.6 Discussion . . . 27

3 Predicting AAA Growth using Graph Convolutional Networks **30**
3.1 Convolutional Neural Networks . . . 30
3.2 Graph Convolutional Networks . . . 31
3.2.1 Learning on 3D surface meshes . . . 31
3.2.2 Gauge-Equivariant Mesh Convolution . . . 33
3.2.3 Pooling in GCNs . . . 33
3.2.4 Network Architecture . . . 34
3.3 Data . . . 34
3.3.1 Surface Meshes . . . 34
3.3.2 Deformation Vectors . . . 35
3.3.3 Input Features . . . 35
3.4 Experiments and Results . . . 36
3.4.1 Time Conditioning . . . 37
3.4.2 Quantitative Evaluation . . . 39
3.5 Discussion . . . 43

4 Discussion and Conclusion **45**
4.1 Strengths and Limitations . . . 45
4.2 Clinical Impact . . . 46
4.3 Future Work . . . 48
4.3.1 ILT Thickness . . . 48
4.3.2 Additional Constraints . . . 48
4.3.3 Latent Space . . . 48
4.3.4 Uncertainty Quantification . . . 49
4.3.5 Clinical Trial . . . 50

4.4 Conclusion 51

References **52**

A Diameter profiles **56**

1 INTRODUCTION

Abdominal aortic aneurysms (AAAs) are focal dilatations of the abdominal aorta that exceed a maximum diameter of 30 mm. They tend to grow and may rupture, which is associated with a mortality rate of approximately 80% [1]. Treatment of AAAs includes either open surgical repair or a minimally invasive alternative known as endovascular aneurysm repair (EVAR). EVAR involves the placement of a stent graft using a catheter that is inserted through the femoral arteries. This stent graft excludes the aneurysm wall from blood circulation, causing the intraluminal thrombus (ILT) to shrink in favorable cases. EVAR is marked by its improvement in peri-operative and early (within six months) morbidity, mortality, and recovery compared to open repair [2, 3]. However, the long term aneurysm-related mortality, re-intervention and rupture rates are higher after EVAR than after open repair [3]. Therefore, whereas patients with a shorter life expectancy are more likely to benefit from EVAR, open repair is recommended for those with prospects of long term survival. Thus, when making recommendations, clinicians should take into account several factors: the risk of AAA rupture, the patient's life expectancy, and the risks of surgical intervention and associated complications. In order to make patient-specific, informed decisions regarding a surgical intervention, it is crucial to carefully assess and compare these factors.

Currently, the assessment of rupture risk primarily relies on the maximum AAA diameter (Figure 1.1, gray steps). Autopsy and clinical studies suggested that the risk of rupture accelerates with increasing AAA diameter [4, 5]. Once diagnosed with AAA, patients are monitored via periodic imaging using various imaging modalities: computed tomography angiography (CTA), magnetic resonance imaging (MRI) or ultrasound (US). If the maximum outer-to-outer wall diameter as measured in these images exceeds a certain threshold (55 mm in men, 50 mm in women) or has grown at a rate of more than 10 mm in a year, surgical treatment is recommended to prevent AAA rupture [6]. These thresholds are based on the results of the United Kingdom Small Aneurysm Trial, a randomised controlled trial of early elective open repair or surveillance for small, symptomless AAAs (40-55 mm). They demonstrated no clear advantage for surgical intervention over surveillance for AAAs that were smaller than 55 mm in diameter [7]. Despite the strong correlation between AAA diameter and rupture risk, several studies show that the maximum AAA diameter is insufficient to assess rupture risk. For example, AAAs with a maximum diameter below the thresholds do rupture while other AAAs with a diameter up to 80 or 90 mm remain stable [8, 9]. This implies that other patient-specific parameters play a substantial role in AAA rupture. Furthermore, since the maximum AAA diameter is measured manually in a certain user-selected slice from a CTA, MRI or US scan, it is associated with high inter-observer variability. In a study by Cayne et al., average measurement variability of maximum AAA diameter on the same CT scan is 4.0 ± 5.1 mm, and can be as much as 35 mm [10]. Here, the CT maximum diameter of 25 AAAs was measured by eight experienced observers. Thus, even when the observers were experienced in issues relating to AAA size, a substantial variability in the measurement of maximum AAA diameter was found. This demonstrates that manual measurement of the maximum AAA diameter on CT scans may not be a true reflection of aneurysm growth.

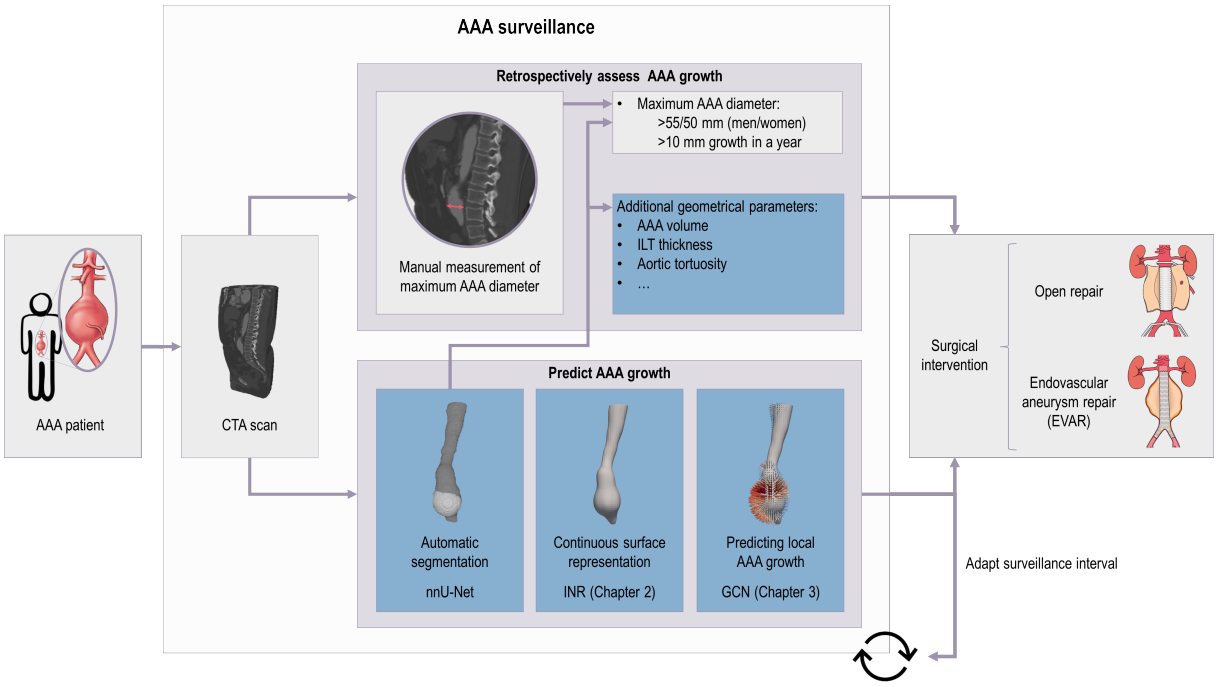


Figure 1.1: In the current pipeline (*gray*), assessment of rupture risk primarily relies on the maximum AAA diameter. In practice, this diameter is measured manually in a certain user-selected slice from a CTA, MRI, or US scan. Our proposed additional steps (*blue*) do not only allow for *retrospectively* assessing AAA growth, but also *predicting* AAA growth over a future time point. Our pipeline consists of the following fully automatic steps: aorta segmentation using an nnU-Net [11], continuous surface representation using an INR [12], and predicting local AAA growth on the surface using a GCN.

Various studies have proposed to use biomechanical parameters, characterizing the complex shape of an AAA, to improve rupture risk assessment [13]. Examples of such parameters are geometrical properties like AAA volume, ILT thickness, and aortic tortuosity, and vessel wall properties like wall shear stress (WSS). These are all based on the aneurysm shape and play a crucial role in the evolution of AAAs. However, manually obtaining the AAA shape is highly time and labor-intensive, requires expertise, and may result in inter and intra-observer variability. Therefore, with an abundance of imaging data available that provide a wealth of valuable information about the aneurysm shape beyond just its maximum diameter, several segmentation algorithms have been developed to automatically extract the AAA shape [14, 15]. In particular, Abdolmanafi et al. even focused on segmentation of different surfaces and volumes comprising the aneurysm, including the wall, lumen, ILT and calcification in the aorta and iliac arteries [15]. This allows for precise, standardized measurement of various biomechanical parameters that may correlate with increased rupture risk.

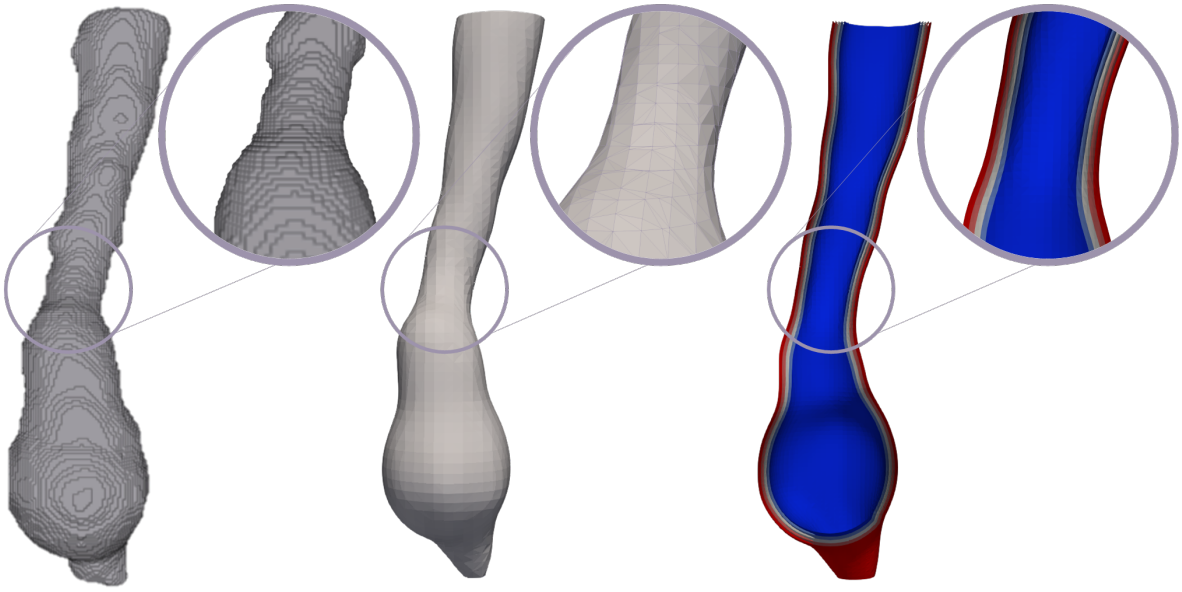
Although these biomechanical parameters are more informative to assess rupture risk than the maximum AAA diameter, these studies do not address the problem of predicting the AAA shape at a future time point. The AAA growth is only considered retrospectively, because a prognostic value for expansion has not yet been acknowledged. In the current pipeline, no prognostic marker for AAA growth and rupture has been implemented as common practice. Prognostic information regarding AAA evolution could have significant benefits in surgical and surveillance management, and provides the patient with valuable information about the expected course of their condition.

Several machine learning techniques have been proposed to predict AAA evolution based on its shape extracted from longitudinal image data. Do et al. utilised a spatio-temporal Gaussian process observation model to construct an implicit surface field, and developed a dynamic model to infer the evolution of this field at a future time [16]. They derive the predicted AAA surface from this predicted field along with uncertainty quantified in future time. Zhang et al. proposed a multistate continuous-time Markov chain model that estimates the transition of AAA progression from one state to another state in terms of AAA growth [17]. Another machine learning model predicts local aortic aneurysm growth based on global, regional, and sub-regional image-derived geometric characteristics, which was proposed by Stoecker et al. [18]. More recently, deep learning methods have been proposed for predicting AAA evolution; Jiang et al. employed a deep belief network to predict AAA expansion [19], and Kim et al. integrated physics-based knowledge with a convolutional neural network by incorporating important multiphysical features [20].

Despite the promising results shown by these methods in predicting AAA evolution, they are often biased by highly sparse and irregularly spaced longitudinal image data. That is, AAA shapes are only represented at the specific time instances when the imaging scans were acquired, providing only snapshots of the AAA shape and associated geometrical parameters at those particular moments. In most studies, data is collected retrospectively, and hence dense and regularly spaced time steps cannot be obtained. For example, in the study by Kim et al. [20], only three longitudinal CTA scans were acquired per patient and their time intervals varied with a minimum of 35 days to a maximum of 769 days. Fixed, regularly spaced time points (one-year) were obtained by linearly interpolating or extrapolating the AAA shape using the second and third CTA scan. This, however, does not take into account the patient-specific AAA growth pattern, and therefore causes a bias in predicting AAA evolution. Furthermore, the quality of the AAA shapes and associated geometrical parameters is often based on the resolution of the image data due to voxel-based representations. Due to this limited availability of smooth AAA shapes at high temporal and spatial resolutions, it is challenging to learn AAA growth rates and predict AAA evolution. To address this challenge, it would be desirable to have a continuous representations of evolving AAA shapes, i.e., at any spatial resolution and any point in time.

For a longitudinal model of an AAA shape, we have to pick a way to represent the 3D shape. There are multiple ways to do this; we distinguish explicit and implicit methods (Figure 1.2). While voxel masks merely discretize an underlying continuous shape and therefore have non-smooth boundaries, mesh-based representations have locally smooth boundaries. However, meshes are still restricted to a fixed resolution, and therefore *explicit* representations. A 3D shape can also be *implicitly* represented by a signed distance function (SDF). The value of an SDF, for a given spatial point, represent the point's distance to the closest point on the surface of the 3D shape; the sign encodes whether the point is inside (negative) or outside (positive) of the surface (respectively blue and red in Figure 1.2). Hence, the zero level set of this SDF represents the underlying continuous surface of the 3D shape (white in Figure 1.2). Since an SDF represents the distance to the surface for any spatial point, it represents the surface with watertight and smooth boundaries at any resolution. Although this way of representing an 3D shape is conceptually appealing due to its continuity, it is difficult to find the actual function. Classical surface reconstruction methods complete a point cloud into a surface by fitting radial basis functions (RBFs) to approximate such an implicit surface function [21].

Recently, implicit neural representations (INRs) have emerged as powerful tools to represent continuous signals on a spatial or spatio-temporal domain [22]. INRs are multilayer perceptrons (MLPs) that take as input continuous coordinates and output the value of the signal or function at that coordinate. The function that is embedded in the MLP can also be an implicit surface



(a) Voxel mask (b) Mesh (c) Signed distance function

Figure 1.2: Different representations of a 3D AAA shape. Both voxel masks (a) and meshes (b) are restricted to a fixed resolution and therefore *explicit* representations. Voxel masks have non-smooth boundaries, whereas boundaries of meshes are locally smooth. A signed distance function (c) *implicitly* represents the surface with smooth and watertight boundaries, at any resolution.

function such as an SDF. By fitting an SDF, INRs are proven capable of encoding surfaces of 3D shapes [23, 24]. This has led to applications in representing 3D shapes evolving over time, like cell shape synthesis [25] and statistical shape modeling [26]. Regarding AAA shapes, Alblas et al. have demonstrated that INRs can be optimized to represent AAA surfaces at a single point in time using a small number of points on the surface [12]. Ideally, you want to extend this model to represent longitudinal AAA surfaces. A continuous representation of an AAA shape evolving over time enables the interpolation of AAA shapes according to the patient-specific growth pattern. This approach addresses the challenge of dealing with highly sparse and irregularly spaced longitudinal image data. Although interpolation allows for retrospectively assessing AAA growth at any moment in-between the scan instances, this approach does not address the prediction of AAA growth over a future time point.

For many years, convolutional neural networks (CNNs) have been a popular deep learning model due to their ability to learn from lots of training data. Hence, they are marked by their generalization capabilities. Although CNNs have shown their many applications for the analysis of medical images, they can only operate on regular Euclidean data (e.g., 2D or 3D grids). Extending CNN models to non-Euclidean domains has been an emerging research area, which is generally referred to as graph convolutional networks (GCNs). Graphs are a type of data structure which consists of a set of objects (vertices) and their relationships (edges). Recent work by Suk et al. has demonstrated the effectiveness of a mesh-based model that uses a GCN on artery surfaces [27]. This method processes signals intrinsically on the artery wall and does not depend on the embedding of local geometry descriptors in 3D. From a technical perspective, predicting AAA growth on the aneurysm surface can be seen as a similar problem. Like [27], we are seeking a 3D vector for each point on the AAA surface. Whereas in the work of [27], this vector represents the WSS, in our work, this vector should represent the growth of the

AAA shape over a future time point. Learning on continuous mesh-based representations of longitudinal AAA surfaces from multiple patients, this GCN can be optimized without any bias from highly sparse and irregularly spaced longitudinal image data. With this, the AAA growth can be predicted for new unseen patients over any desired future time point.

In this work, I propose an additional, automatic pipeline that does not only allow for *retrospectively* assessing AAA growth, but also *predicting* AAA growth over a future time point (Figure 1.1, blue steps). Using INRs, I aim to obtain highly personalized, continuous representations of AAA surfaces evolving over time, based on sparse and irregularly spaced longitudinal CTA data (Chapter 2). In particular, I investigate to what extent such a model can be used to inter- and extrapolate AAA surfaces in time for different temporal regularizations. Additionally, I propose to use a GCN to predict AAA growth over a future time point, based on the longitudinal representations from multiple patients (Chapter 3). In Chapter 4, I will discuss the strengths and limitations of these models and their potential to aid physicians in clinical decision-making.

Thus, the objective of this study is to address the following main research question, with subquestions for Chapter 2 and 3:

- How can deep learning be used to predict personalized aneurysm growth based on longitudinal image data?
  - Chapter 2: To what extent can implicit neural representations (INRs) be used to inter- and extrapolate AAA surfaces in time, based on sparse and irregularly spaced longitudinal CTA data?
  - Chapter 3: To what extent can a graph convolutional network (GCN) predict local AAA growth, based on longitudinal aneurysm representations from multiple patients?

2 MODELING AAA EVOLUTION USING IMPLICIT NEURAL REPRESENTATIONS

In this chapter, I demonstrate how we model AAA evolution based on highly sparse and irregularly spaced longitudinal CTA data. In particular, I show how we obtain continuous representations of AAA surfaces evolving over time, at any spatial resolution and any point in time. I investigate to what extent this model can be used to inter- and extrapolate AAA surfaces in time.

2.1 Shape Representations

Before delving into representing an AAA shape evolving over time, we have to pick a way to represent a 3D shape at a single point in time. We distinguish *explicit* methods such as voxel masks and meshes, and *implicit* methods such as a signed distance function. In Figure 1.2, we have shown these representations for a 3D AAA shape.

Voxel mask A voxel mask represents a 3D shape on a Euclidean domain, namely a 3D grid of small volumetric elements called voxels. It is a binary discretization of an underlying continuous shape. This explicit representation has non-smooth boundaries and a fixed resolution. Voxel masks have been a popular representation because medical image data are represented in the same way, namely pixels or voxels. Convolutional neural network-based methods have shown their effectiveness in dealing with this type of representation by semantic segmentation of various shapes from image data. Hence, the quality of voxel masks is typically dependent on the resolution of the image data.

Mesh A mesh consists of vertices, edges and faces that represent a 3D shape. Its boundaries are locally smooth, but they are still restricted to a fixed resolution and therefore explicit representations.

Signed distance function A shape can be implicitly represented by a signed distance function (SDF). In the case of a 2D surface $\mathcal{M}$ embedded in a 3D domain, $SDF_{\mathcal{M}}(\boldsymbol{x}) : \mathbb{R}^3 \mapsto \mathbb{R}$ is defined as:

$$SDF_{\mathcal{M}}(\boldsymbol{x}) = \begin{cases} -d(\boldsymbol{x}, \mathcal{M}) & \boldsymbol{x} \text{ inside } \mathcal{M} \\ 0 & \boldsymbol{x} \text{ on } \mathcal{M} \\ d(\boldsymbol{x}, \mathcal{M}) & \boldsymbol{x} \text{ outside } \mathcal{M}. \end{cases} \tag{2.1}$$

The value of $d(\boldsymbol{x}, \mathcal{M})$ represent the distance to the closest point on the surface $\mathcal{M}$, for a given spatial point $\boldsymbol{x}$. The sign of the SDF encodes whether the point is inside (negative) or outside (positive) of the surface. Hence, the zero level set of this SDF represents the underlying shape's surface. Since an SDF outputs the distance to the surface for any spatial point, it represents the surface with watertight and smooth boundaries at any resolution.

An implicit representation of a 3D shape can be easily transformed into any explicit representation due to its continuity, while the reverse process is a more complex process that requires techniques such as surface reconstruction. Classical surface reconstruction methods complete a point cloud into an implicit surface by fitting RBFs. Furthermore, conversion between explicit representations is a challenging task and also requires reconstruction techniques. An example is marching cubes, which is an algorithm that extracts a surface mesh from a 3D voxel mask.

2.2 Time-dependent Signed Distance Function

In this work, we are interested in modeling AAA evolution. Thus, instead of representing a single AAA shape, we want to represent an AAA shape evolving over time. While a single shape is described by its spatial coordinates, a shape evolving over time is also described by a time coordinate. Regarding AAA evolution, we encounter a challenge in terms of temporal resolution. That is, available longitudinal image data is typically highly sparse and irregularly spaced in time. Therefore, we are seeking a representation that is able to fill in the gaps in this surveillance data, i.e., interpolate in-between the scan instances. An implicit representation of evolving AAA shapes is well-suited for this task.

A surface evolving over time can be implicitly represented by the zero level set of its *time-dependent* SDF. In this work, we consider a 2D surface $\mathcal{M}$ of a shape evolving over time, embedded in a 4D domain. Its time-dependent $SDF_{\mathcal{M}}(\boldsymbol{x}, t) : \mathbb{R}^3 \times \mathbb{R} \mapsto \mathbb{R}$ is defined as:

$$SDF_{\mathcal{M}}(\boldsymbol{x}, t) = \begin{cases} -d(\boldsymbol{x}, \mathcal{M}) & \boldsymbol{x} \text{ inside } \mathcal{M} \text{ at time } t \\ 0 & \boldsymbol{x} \text{ on } \mathcal{M} \text{ at time } t \\ d(\boldsymbol{x}, \mathcal{M}) & \boldsymbol{x} \text{ outside } \mathcal{M} \text{ at time } t. \end{cases} \tag{2.2}$$

The value of $d(\boldsymbol{x}, \mathcal{M})$ represents the minimum distance to the surface $\mathcal{M}$ at position $\boldsymbol{x}$ at time point t. Hence, the zero level set of this SDF represents the surface $\mathcal{M}$. A time-dependent SDF should satisfy the Eikonal equation at each time point, hence $||\nabla_{\boldsymbol{x}} SDF_{\mathcal{M}}(\boldsymbol{x}, t)|| = 1, \forall \boldsymbol{x}, \forall t$.

In theory, this time-dependent SDF provides a continuous representation of a surface evolving over time. This is conceptually appealing as it maintains continuity throughout the evolution of the surface at any spatial and temporal resolution. However, in practice, it is difficult to find the actual function. To this end, we use a neural network that we can optimize to represent the time-dependent SDF of a patient-specific AAA surface evolving over time.

2.3 Implicit Neural Representations

In previous work, it has been shown that an SDF of a fixed or evolving surface can be represented by a neural network [23, 24, 25, 26, 12]. In this work, we embed the SDF of an AAA surface evolving over time in a so-called *implicit neural representation* (INR). Our INR is a fully connected multilayer perceptron (MLP), $f(\boldsymbol{x}, t; \theta)$, that takes as input a 4D coordinate from the spatio-temporal domain $\Omega := [-1, 1]^3 \times [-1, 1]$, and outputs $SDF_{\mathcal{M}}(\boldsymbol{x}, t)$. Figure 2.1 shows a schematic overview of our INR. As the INR is trained on continuous spatio-temporal coordinates, it allows for an AAA representation at any spatial resolution and any point in time.

2.3.1 Optimizing an INR

Our network can be optimized to represent an AAA surface evolving over time, based on longitudinal image data. In particular, we train our network using a sequence of J input point clouds $\{\mathcal{X}_{t_j}\}_{j=1,\dots,J}$, where $\mathcal{X}_{t_j} = \{\boldsymbol{x}_i\}_{i \in I} \subset [-1, 1]^3$ represents points on the AAA surface of a single

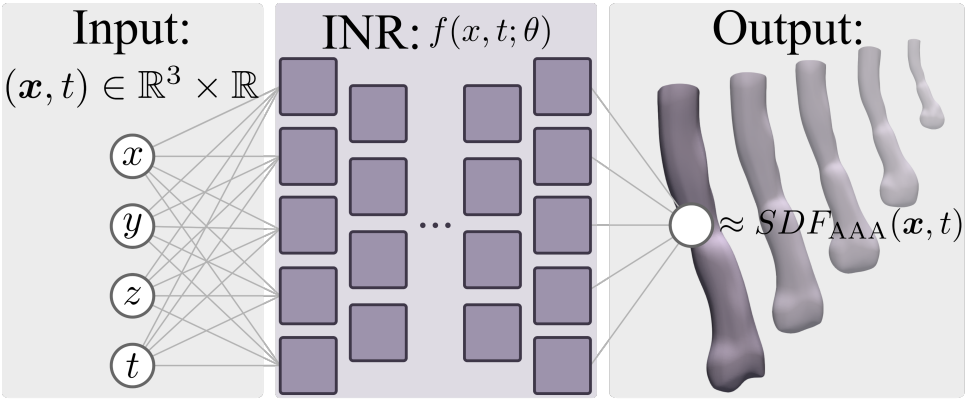


Figure 2.1: Schematic representation of our INR, taking spatio-temporal coordinates $(\boldsymbol{x}, t)$ as input, outputting $SDF_{\mathcal{M}}(\boldsymbol{x}, t)$ of the AAA surface. Note that a single INR represents the complete evolving AAA surface of a single patient. Copied from [28].

patient at time t_j. Note that J denotes the number of scan instances for a single patient. Using these point clouds, we optimize the parameters θ in our MLP $f(\boldsymbol{x}, t; \theta)$ so that it approximates $SDF_{\mathcal{M}}(\boldsymbol{x}, t)$ to a plausible AAA surface evolving over time.

The loss function we use to optimize the INR consists of three terms: a data term $\mathcal{L}_{\mathcal{X}_{t_j}}$, and two regularization terms $\mathcal{L}_{\text{Eikonal}}$ and $\mathcal{L}_{\text{temporal}}$:

$$\mathcal{L}(\theta) = \sum_{1 \leq t_j \leq J} (\mathcal{L}_{\mathcal{X}_{t_j}}(\theta) + \lambda_1 \mathcal{L}_{\text{Eikonal}}(\theta)) + \lambda_2 \mathcal{L}_{\text{temporal}}(\theta), \tag{2.3}$$

where λ_1 and λ_2 are hyperparameters.

The data term $\mathcal{L}_{\mathcal{X}_{t_j}}$ was previously introduced in [24] and ensures that points that are known to be on the AAA surface are indeed in the zero level set of the time-dependent SDF. For each time t_j, we sample I spatial points from the input point cloud $\mathcal{X}_{t_j}$, and encourage $f(\boldsymbol{x}, t_j; \theta)$ to vanish on these points. Thus, $\mathcal{L}_{\mathcal{X}_{t_j}}$ is defined as:

$$\mathcal{L}_{\mathcal{X}_{t_j}}(\theta) = \frac{1}{|I|} \sum_{i \in I} (|f(\boldsymbol{x}_i, t_j; \theta)|). \tag{2.4}$$

The regularization term $\mathcal{L}_{\text{Eikonal}}$ was also previously introduced in [24] and ensures that the INR represents a proper SDF. For each time t_j, we now sample M points from the spatial domain $\Omega := [-1, 1]^3$, and encourage that $f(\boldsymbol{x}, t_j; \theta)$ is a solution to the Eikonal equation. Thus, $\mathcal{L}_{\text{Eikonal}}$ is defined as:

$$\mathcal{L}_{\text{Eikonal}}(\theta) = \mathbb{E}_{\boldsymbol{x}}(||\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, t_j; \theta)|| - 1)^2. \tag{2.5}$$

The regularization term $\mathcal{L}_{\text{temporal}}$ was recently introduced in [28] to avoid inconsistent representations of the AAA surface for time points where point cloud data is unavailable, i.e. in-between the scan instances. For this we sample N points from the spatio-temporal domain $\Omega := [-1, 1]^3 \times [-1, 1]$, and regularize $f(\boldsymbol{x}, t; \theta)$ at times where the AAA surface is unknown. In this work, we evaluate two temporal regularization strategies.

Restricting any temporal change We restrict any temporal change by minimizing the temporal gradient of the SDF. In this case, $\mathcal{L}_{\text{temporal}}$ is defined as:

$$\mathcal{L}_{\text{temporal}}(\theta) = \mathbb{E}_t(||\nabla_t f(\boldsymbol{x}, t; \theta)||). \tag{2.6}$$

Allowing growth but not shrinkage During pre-operative AAA surveillance, AAAs typically tend to grow but not shrink. We can use this to impose more plausible constraints on the time-dependent SDF. Note that the SDF is negative on the inside and positive on the outside of the AAA surface, and therefore the temporal gradient is negative for an AAA surface that grows. Thus, by setting $\lambda_2 > 0$ and minimizing only positive values of the temporal gradient, we allow the AAA surface to grow but not to shrink. In this case, $\mathcal{L}_{\text{temporal}}$ is defined as:

$$\mathcal{L}_{\text{temporal}}(\theta) = g(\mathbb{E}_t(\nabla_t f(\boldsymbol{x}, t; \theta))), \tag{2.7}$$

where $g(x) = max(0, x)$ is a ReLU activation function.

2.4 Data

To evaluate the INR's performance in representing a surface evolving over time, we use two kinds of datasets. We create a *synthetic* dataset, consisting of toy surfaces with different growth patterns. In this dataset, we have control over the number of time points and their intervals, i.e., how dense and regularly spaced the data is. Furthermore, we include an *aneurysm* dataset that is based on longitudinal CTA data from multiple AAA patients. It is important to note that this dataset introduces confounding factors that may affect the INR's performance in representing an evolving surface, which will be addressed in Section 2.4.2.

2.4.1 Synthetic Dataset

We created three longitudinal synthetic datasets, each consisting of 11 spheres, with different growth patterns in terms of diameters: exponential growth, linear growth, and linear shrinkage. We chose spheres as toy surfaces because their growth pattern is the same for each point on the surface. This way, we do not have any difference across the surface regarding local growth, and therefore the growth pattern is known for the whole surface. The diameters of the spheres were based on real-world AAA diameters, and ranged from 32 to 85 mm. Each sphere was created using the icosphere function of Trimesh [29], and is described by a tuple of vertices and faces $\mathcal{M} = (\mathcal{V}, \mathcal{F})$. For each sequence of 11 spheres $\{\mathcal{M}_{t_j}\}_{j=0,\dots,10}$, the set of vertices $\{\mathcal{V}_{t_j}\}_{j=0,\dots,10}$ were jointly normalized to the $[-1, 1]^3$ domain. The time points t_0-t_{10} were equally spaced and normalized to the $[-1, 1]$ interval. The resulting synthetic dataset is shown in Figure 2.2 and can be used as input for the INR.

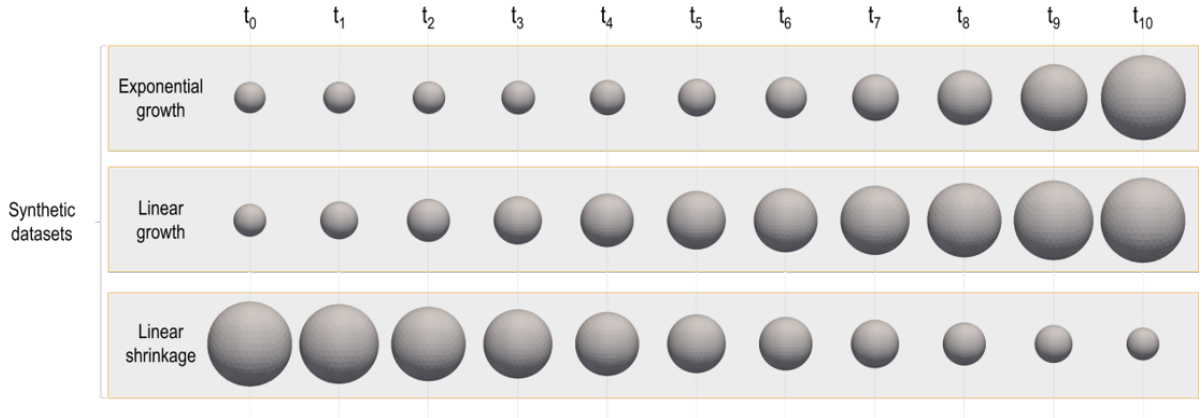


Figure 2.2: Longitudinal synthetic datasets, each consisting of 11 spheres, with different growth patterns: exponential growth (*top*), linear growth (*middle*), and linear shrinkage (*bottom*).

2.4.2 Aneurysm Dataset

Study Subjects

We retrospectively included longitudinal CTA scans of 25 AAA patients collected at Seoul National University Hospital (Seoul, South Korea) from 2000 to 2016. This longitudinal image data was kindly shared with us by the authors of [16, 17, 19, 20]. The study population consisted of 24 males and 1 female, with their age at the first scan instance ranging from 54 to 80 years. In total, 117 CTA scans were acquired using a CT scanner. The number of CTA scans per patient ranged from 2 to 7, with time intervals between consecutive CTA scans ranging from 11 to 1713 days, meaning that we have highly irregularly spaced, sparse data at fixed time points. The CTA scans had an in-plane resolution ranging from 0.38 to 0.87 mm and an axial (z-axis) resolution ranging from 0.70 to 3.00 mm, indicating some diversity in image quality. In Table 2.1, the details about the subject data are summarized.

Table 2.1:

Summary of Subject Data	
Gender (male/female)	24/1
Age at first scan instance (years)	67 (54-80)
Total CTA scans	117
Number of CTA scans per patient	5 (2-7)
Time interval between CTA scans (days)	338 (11-1713)
In-plane resolution (mm)	0.38-0.87
Axial (z-axis) resolution (mm)	0.70-3.00

Lumen and Thrombus Segmentation

We obtained automatic segmentations of the lumen and thrombus of the abdominal aorta in each CTA scan using an nnU-Net [11] trained on 80 pre-operative CTA scans of AAA patients with abdominal aorta annotations. This training dataset was acquired at Amsterdam AMC (Amsterdam, The Netherlands) using a CT scanner. The in-plane resolution ranged from 0.63 to 0.98 mm, and the axial (z-axis) resolution ranged from 0.50 to 2.00 mm. The annotation protocol of the training dataset included two steps: 1) lumen and thrombus point clouds were annotated in the pre-operative CTA scans, and 2) surface reconstructions of these manually annotated point clouds were made using an INR [12]. Quantitative evaluation of the trained nnU-Net model in a separate test dataset consisting of 13 CTA scans showed a median Dice similarity coefficient (DSC) of 0.90 ± 0.09 for AAA segmentation (including both lumen and thrombus).

Using this trained nnU-Net, we successfully obtained automatic lumen and thrombus segmentations in our dataset. However, six lumen segmentations of five patients required manual correction using 3D Slicer [30] as they had not been properly segmented by the trained nnU-Net. This is most likely due to the out-of-distribution nature of our dataset, which includes factors such as the use of a different CT scanner and acquisition protocol compared to the training data of the nnU-Net. In addition, our dataset consists of patients from South Korea, whereas the nnU-Net training data primarily included patients from The Netherlands. The obtained lumen and thrombus segmentation together form the aorta segmentation.

From the aorta segmentations we can automatically extract the maximum diameter based on the maximum inscribed sphere (MIS) method [31, 19]. In Figure 2.3, we show the maximum diameter for each patient over time based on age and scan instances.

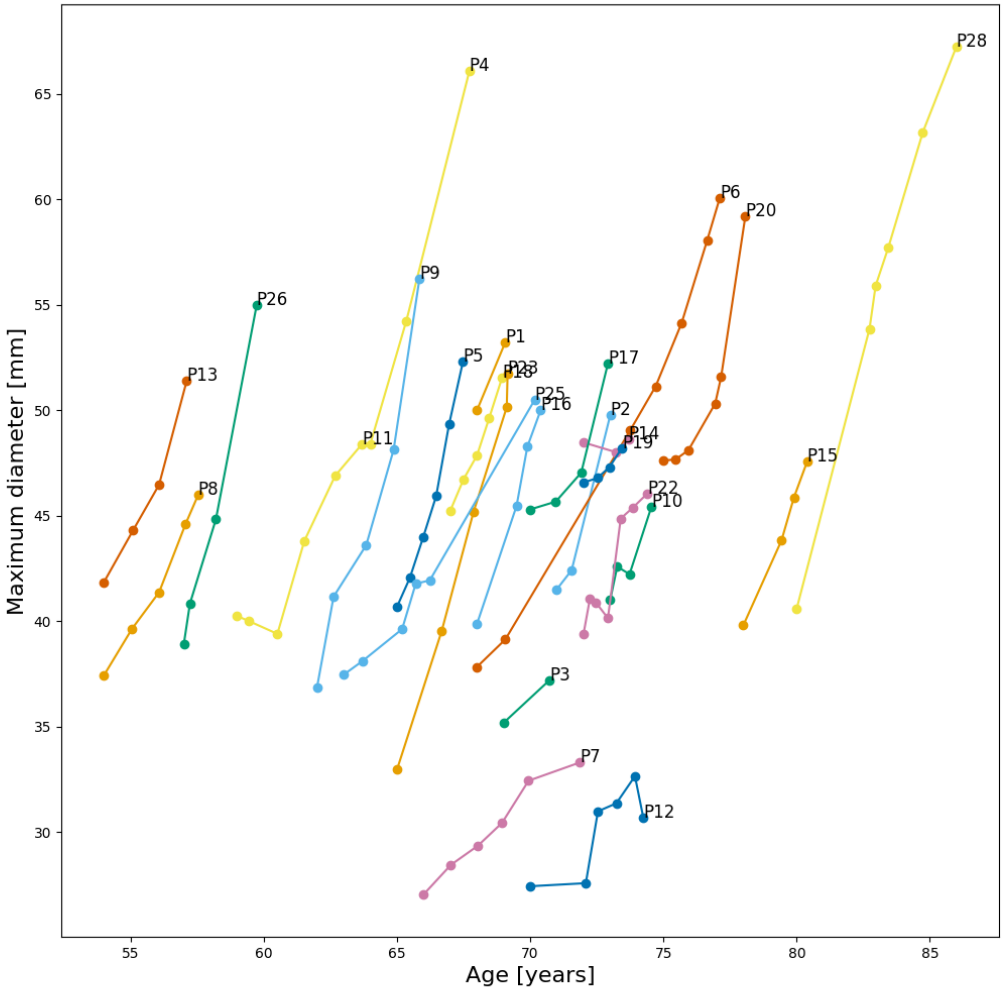


Figure 2.3: Automatically extracted, maximum inscribed sphere (MIS) diameters from the abdominal aortic aneurysm (AAA) segmentations for each patient over time based on age and scan instances in years. Each line corresponds to one patient and the dots indicate the patient's age at which the computed tomography angiography (CTA) scan was taken.

Rigid Registration

In order to model local AAA evolution, it is crucial that all AAA shapes of the same patient are aligned in the same coordinate system. For this, we used rigid registration within the Insight ToolKit (ITK) image registration framework [32], based on the aorta segmentations. Note that we do not want more degrees of freedom like scaling to register the AAA shapes. That is, we only aim to align the AAA shapes because the INR model should learn its local evolution.

Aorta Point Clouds

The surface of each registered aorta segmentation was extracted and represented as a point cloud. For each patient, the spatial coordinates of all registered point clouds were jointly nor-

malized to the $[-1, 1]^3$ domain, as for the synthetic dataset. Similarly, the time points of the scan instances were normalized to the $[-1, 1]$ interval. The resulting sequence of normalized point clouds can then be used as input for the INR.

Confounding Factors

In this aneurysm dataset, several confounding factors can affect the INR's performance in representing an AAA surface evolving over time.

- **Number of scan instances and their time intervals.** We have highly irregularly spaced, sparse data at fixed time points. It is expected that the INR can better model AAA evolution by learning a growth pattern from dense and regularly spaced data. We, however, do not have any control over the number of scan instances and their time intervals because the data is collected retrospectively.
- **Image quality.** There are several factors that affect the image quality and thereby the representation of the AAA shape. Examples are insufficient contrast enhancement of the lumen, limited in-plane or axial resolution, inadequate field of view (FOV), and CT settings like peak kilovoltage (kVp).
- **Segmentation.** The AAA segmentations were obtained automatically using a trained nnU-Net. The training data for this nnU-Net, however, differs in terms of distribution compared to our dataset. For example, patients from the nnU-Net training data came from The Netherlands, whereas patients from our dataset come from South-Korea. This may have caused some segmentations errors in our dataset.
- **Registration.** If the AAA shapes are not registered perfectly to each other, the INR may be prone to learn the global translation and rotation to better align AAA shapes, instead of learning local growth of the aneurysm.
- **Local growth.** AAA growth is typically assessed in terms of maximum diameter. As shown in Figure 2.3, we can obtain a global growth trend by fitting the maximum diameters. Global growth, however, is insufficient to describe the AAA evolution. Whereas the dilated part of the AAA shape grows, healthy parts remain stable. Therefore, local growth across the AAA surface varies, which may result in distinct performance of the INR for different surface regions. But since local growth is unknown, we cannot evaluate this.

It is important to note that in our synthetic dataset, we either eliminated these confounding factors or had control over them. In particular, we are free of diversity in image quality, and segmentation and registration errors, and have control over the temporal resolution and local growth of the evolving toy surface.

2.5 Experiments and Results

To represent a shape evolving over time, the INRs consisted of an MLP with three fully connected layers and a single final node. For the synthetic dataset, we used 128 nodes for each layer with Sine activation functions ($\omega = 2$), resulting in a SIREN model [33]. For the aneurysm dataset, we used 256 nodes for each layer with Softplus activation functions ($\beta = 100$). Like [24, 23], we used a single skip connection from the input to the middle layer. Note that we used a different activation function for the aneurysm dataset than for the synthetic dataset. This is because the INR is highly sensitive to the hyperparameter ω, and requires tuning for each different shape. For the toy surfaces, we only have to tune this parameter ones, but for aneurysm shapes, we should tune this parameter for each patient. We set our loss weights λ_1 and λ_2

Table 2.2: Hyperparameter settings

Hyperparameters	Synthetic dataset	Aneurysm dataset
Number of layers	3	3
Number of nodes per layer	128	256
Activation function σ	Sine ($\omega = 2$)	Softplus ($\beta = 100$)
Loss weights	$\lambda_1 = 0.1$ No temporal regularization: $\lambda_2 = 0$ With temporal regularization: $\lambda_2 = 0.1$	$\lambda_1 = 0.1$ No temporal regularization: $\lambda_2 = 0$ With temporal regularization: $\lambda_2 = 0.05$ or $\lambda_2 = 0.1$
Optimizer	Adam	Adam
Learning rate	10^{-5}	10^{-4}
Number of epochs	10000	25000
Batch size N	300	6000

(Equation 2.3) as shown in Table 2.2. We used an Adam optimizer with a constant learning rate to train our network on an NVIDIA Quadro RTX 6000 GPU. In Table 2.2, the learning rate and number of epochs are shown for the synthetic and aneurysm dataset. In each epoch, we randomly sampled the following coordinates for each time point t_j:

- N points from the input point cloud $\mathcal{X}_{t_j} \subset [-1, 1]^3$ with time point $t_j \subset [-1, 1]$,
- N points from spatial domain $\Omega := [-1, 1]^3$ with time point $t_j \subset [-1, 1]$, and
- N points from spatio-temporal domain $\Omega := [-1, 1]^3 \times [-1, 1]$,

where N is the batch size (Table 2.2). Note that we use different hyperparameters for the aneurysm dataset than for the synthetic dataset. For example, the number of nodes per layer, the number of epochs and the batch size are much larger. This is because representing real-world evolving AAA shapes is a more challenging task for an INR compared to representing spheres that evolve over time based on a well-defined growth pattern.

In this work, we evaluate how different temporal regularization strategies affect the INR's performance in representing an evolving shape. In particular, we compare two temporal regularization strategies: restricting any temporal change by minimizing the temporal gradient of the SDF (TempReg), and allowing growth but not shrinkage by minimizing only positive values of the temporal gradient of the SDF (ReLUTempReg). We also evaluate the INR's performance without applying any temporal regularization (NoTempReg). That is, by setting $\lambda_2 = 0$, we do not impose any temporal constraints on the SDF.

2.5.1 Continuous Sphere Representation

For each of the three longitudinal synthetic datasets, we optimized a single INR based on the set of vertices $\{\mathcal{V}_{t_j}\}_{j=0,\dots,10}$ at all time points, which will be referred to as the reference experiment. We use this fully sampled INR to extract the surfaces of the toy spheres at the ground-truth time points $t_{j=0,\dots,10}$ and their intermediate time points $t_{j=0.5,\dots,9.5}$. Note that since the INR is trained on continuous spatio-temporal coordinates, we can extract a sphere surface at any spatial resolution and any point in time. For each extracted sphere surface, we compute the diameter by taking the mean distance of all its vertices $\mathcal{V}$ to the centre and multiplying this radius with two. In the *top row* in Figure 2.4, we compare the diameters of the ground-truth and extracted spheres for the reference experiments for different growth patterns and temporal regularization strategies.

We observe that the INRs can accurately *interpolate* the sphere's surface in time. That is, the diameters of the intermediate time points precisely follow the ground-truth growth patterns. This

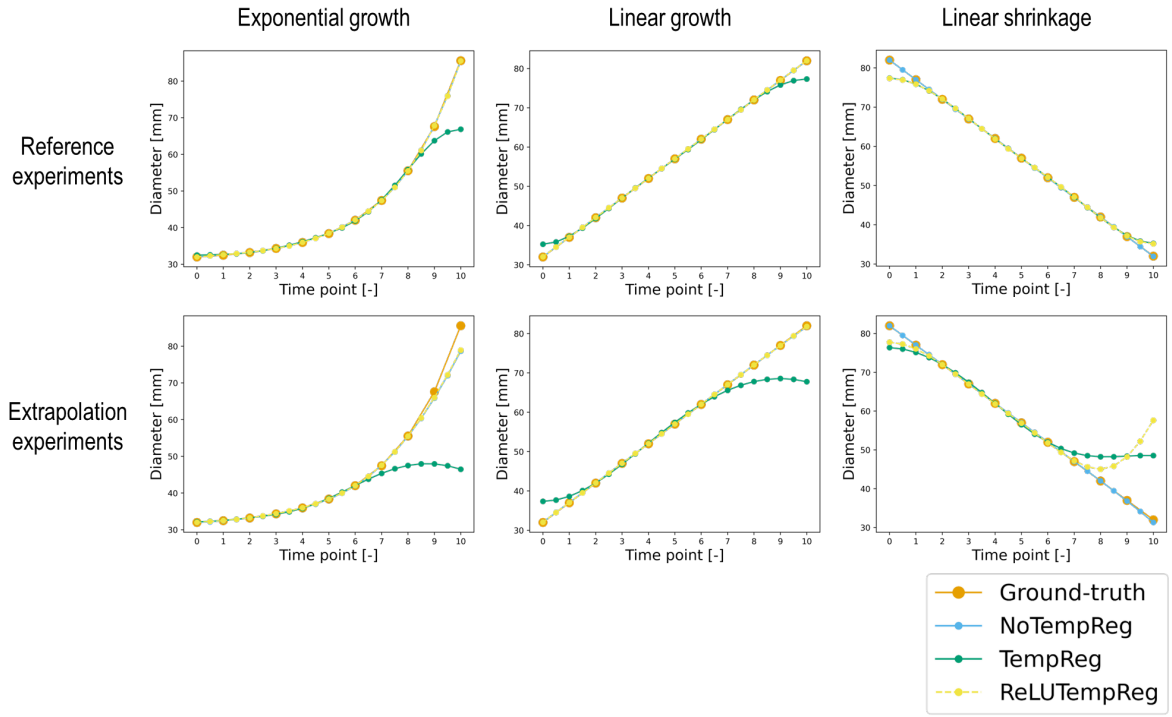


Figure 2.4: Diameters of ground-truth and extracted spheres for the reference and extrapolation experiments (*top* and *bottom row*, respectively).

indicates that the INR is able to learn a well-defined growth pattern and represent a sphere surface evolving over time at any temporal resolution.

We compare how different temporal regularization strategies (NoTempReg, TempReg, and ReLUTempReg) affect the INR's performance in representing the sphere's surface evolving over time. For both exponential and linear growth (*top left* and *middle* in Figure 2.4), the diameters from the INR optimized with NoTempReg (*blue*) and ReLUTempReg (*yellow*) precisely follow the ground-truth diameters. For TempReg (*green*), however, the diameters flatten at both ends of the time points that were seen during training. This is the case for all growth patterns and is most likely due to regularizing any temporal change. Interestingly, the flattening is also observed for ReLUTempReg when the surface of the sphere shrinks over time (*top right* in Figure 2.4). This indicates that this temporal regularization performs as expected; it allows growth but restricts shrinkage. Thus, according to these reference experiments, optimizing an INR with no temporal regularization (NoTempReg) results in the most accurate, continuous representation of the sphere's surface evolving over time.

Extrapolation

In order to evaluate the INR's performance in *extrapolating* the sphere's surface over time, we optimized an INR similarly as the reference experiment but left out the last two time points (t_9 and t_{10}) during training. Thus, for each growth pattern, we optimize a single INR based on the set of vertices $\{\mathcal{V}_{t_j}\}_{j=0,\dots,8}$. Using this optimized INR, we reconstruct the surfaces of the toy spheres at the left-out time points t_9 and t_{10} and compare it to the ground-truth spheres. In the *bottom row* in Figure 2.4, we compare the diameters of the ground-truth and extracted spheres for these extrapolation experiments. In addition to these diameter plots, we show the corresponding extracted spheres in Figure 2.5.

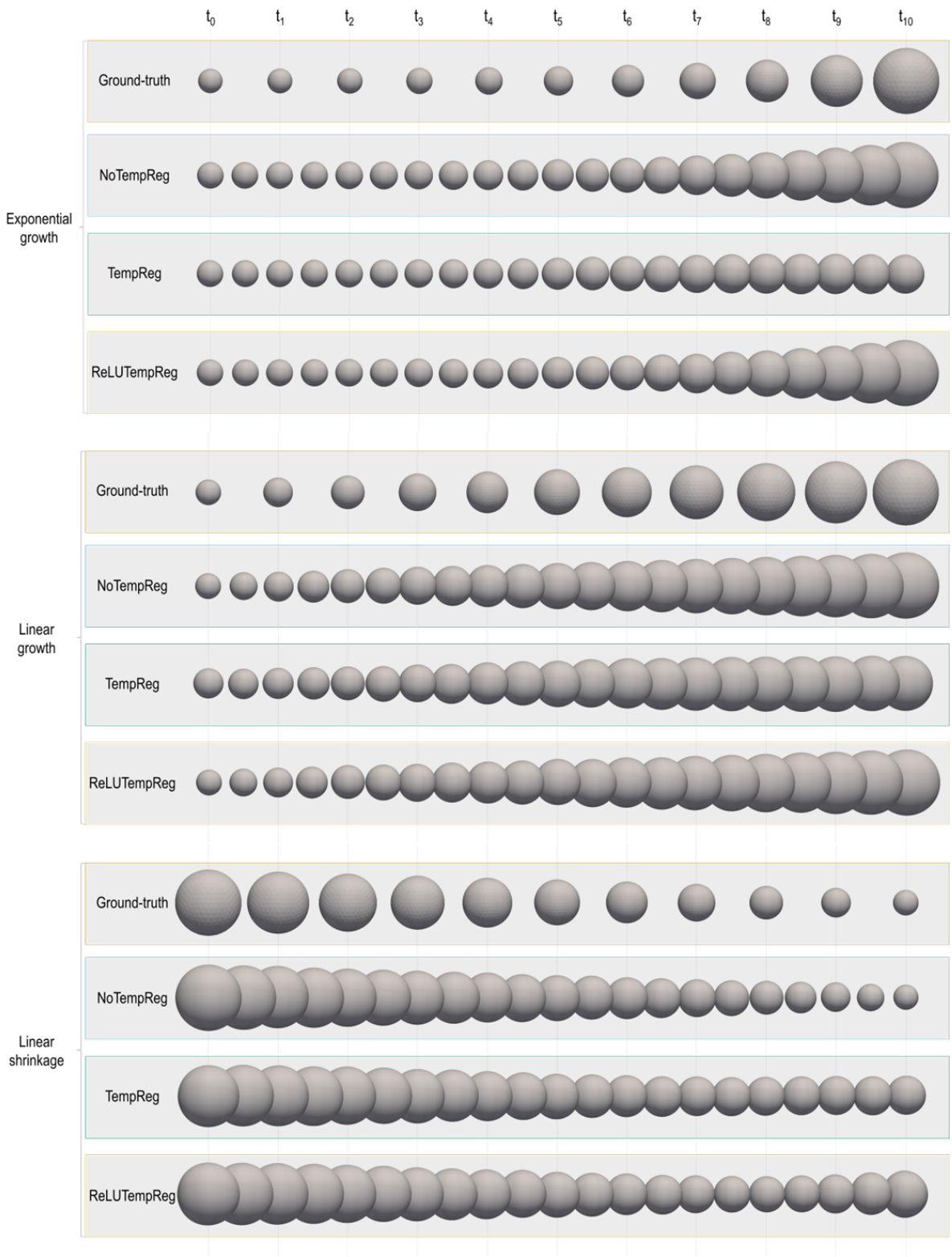


Figure 2.5: Ground-truth and extracted spheres for the extrapolation experiments. For each growth pattern (*top*, *middle*, and *bottom*), we compare the results for different temporal regularization strategies: NoTempReg (*blue*), TempReg (*green*), and ReLUTempReg (*yellow*). Note that the INRs were optimized based on the ground-truth spheres from time points t_0 to t_8. Hence, extracted spheres at intermediate time points are interpolations, and after time point t_8 are extrapolations.

For both exponential and linear growth (*bottom left* and *middle* in Figure 2.4), optimizing the INR with with NoTempReg (*blue*) and ReLUTempReg (*yellow*) resulted in the best extrapolation performance. Whereas the INR was able to learn the linear growth pattern and accurately extrapolate spheres, it was not able to fully learn the exponential growth pattern. As for the reference experiments, we observe that for TempReg (*green*) the diameters flatten at both ends of the time points that were seen during training. Note that now the flattening occurs earlier because the last two scans were left out during training. Interesting behavior is observed for ReLUTempReg when the surface of the sphere shrinks over time (*bottom right* in Figure 2.4). That is, the sphere's surface starts to grow after t_8, which is the last time point seen during training. It seems that the INR tends to predict growth rather than learning from previous time points that the sphere should shrink. Thus, according to these extrapolation experiments, optimizing an INR with no temporal regularization (NoTempReg) results in the best extrapolation performance.

In conclusion, according to this toy problem, we obtained the following key observation:

- **Temporal regularization is not necessary for an INR to represent plausible surfaces for time points where data is unavailable, i.e. in-between the training time points.**
- **An INR model that is optimized with no temporal regularization is able to accurately learn a growth pattern and even extrapolate on this growth pattern.** With temporal regularization, the growth pattern tends to flatten at both ends of the training time points, which makes it difficult to accurately extrapolate on this growth pattern.

Now that we know how different temporal regularization strategies affect the INR's performance in representing and inter- and extrapolating an evolving shape, we will investigate the same for our aneurysm dataset. Note that now the confounding factors that were mentioned earlier come into play.

2.5.2 Continuous AAA Representation

In this section, we show the results for five patients with their number of available scan instances ranging from 3 to 7, namely P2, P26, P9, P7, and P6 from the aneurysm dataset (Figure 2.3). For clarity, we will refer to them as Patient 1, Patient 2, Patient 3, Patient 4, and Patient 5, respectively.

For each patient, we optimized a single INR based on point clouds from all available scan instances, which will be referred to as the reference experiment. As the INR is trained on continuous spatio-temporal coordinates, we can extract an AAA shape at any resolution and any point in time. In Figure 2.6, we show 11 interpolated AAA shapes of five patients at regularly spaced time intervals, extracted from the fully sampled INRs optimized with TempReg. In Figure 2.6b, we compare the diameter profiles along the AAA centerline of the ground-truth aorta segmentations (*solid lines*) to the AAA surfaces extracted by the optimized INR (*dashed lines*). We observe that these fully sampled INRs closely resemble the AAA shapes at the scan instances. Thus, according to this reference experiment, optimizing the INR with TempReg results in accurate, continuous representations of the AAA shapes evolving over time. Therefore, we use these representations at the scan instances as a reference to evaluate the next experiment.

In Figure A.1 in Appendix A, we show diameter profiles along the AAA centerline for the ground-truth aorta segmentations (*black dashed lines*) and interpolated AAA shapes extracted at time points spaced half a year apart (*colored solid lines*) for all 25 included patients.

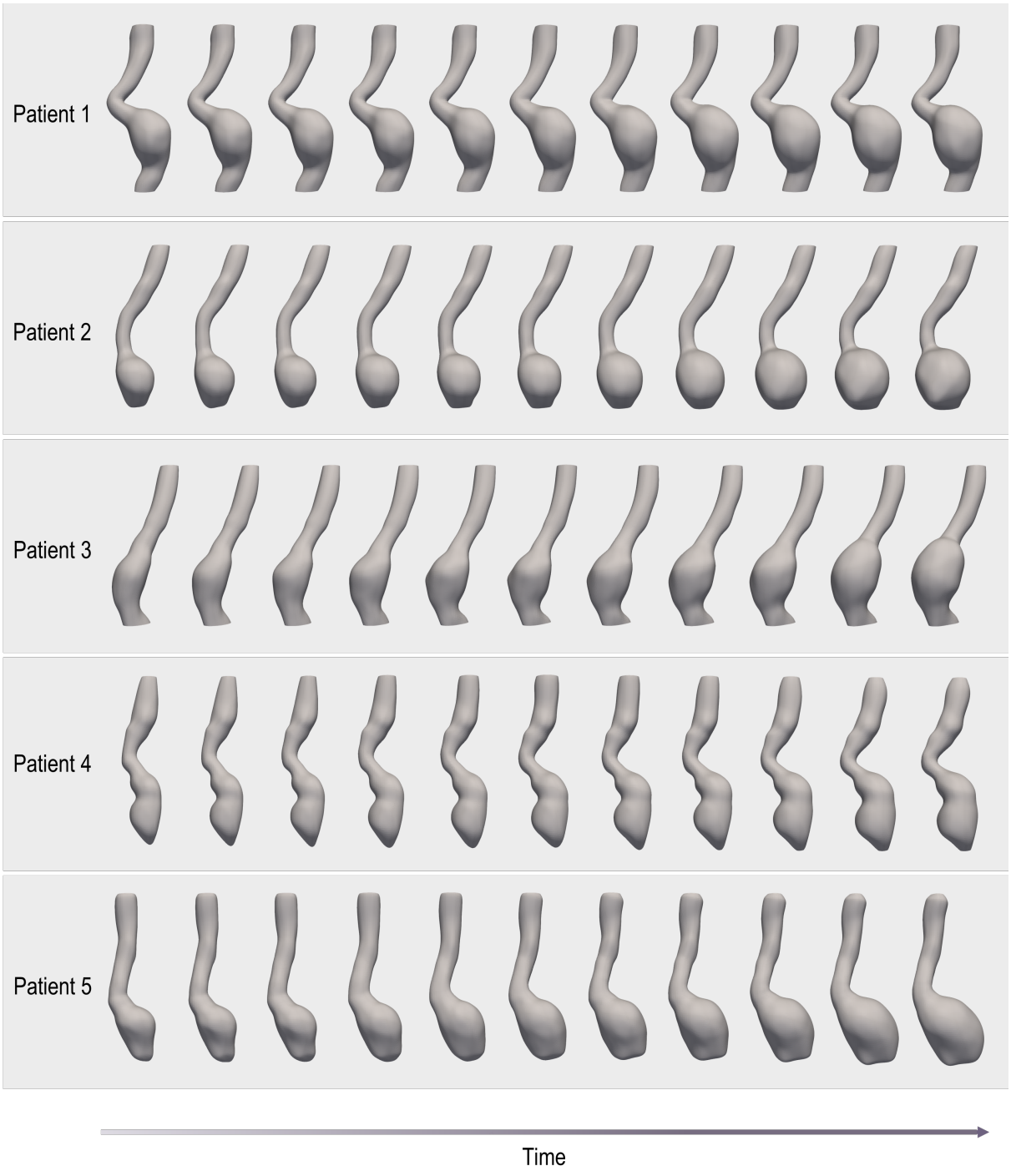


Figure 2.6: Interpolated AAA shapes at regularly spaced time intervals for five patients. The AAA shapes are extracted from the fully sampled INRs optimized with TempReg.

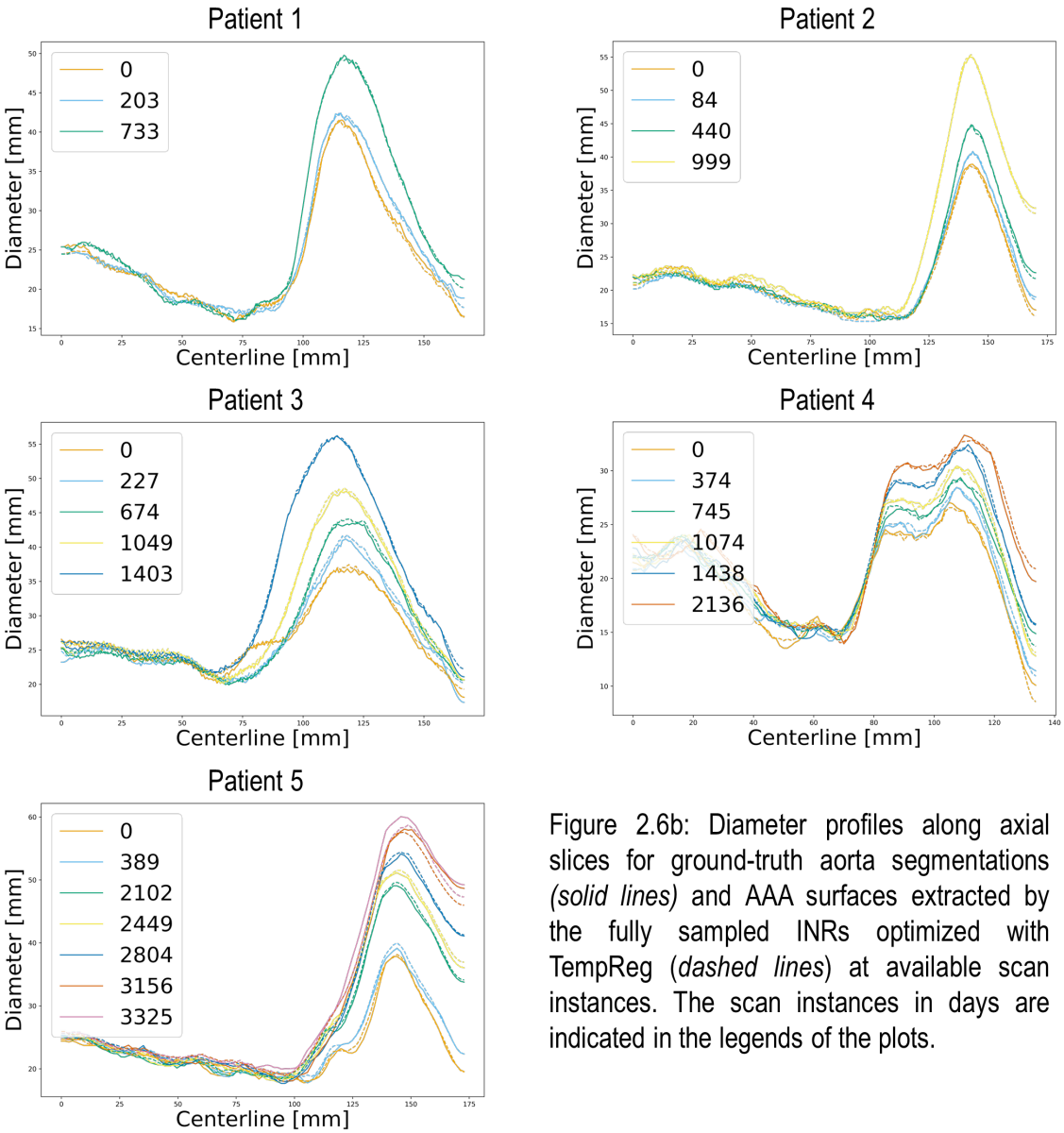


Figure 2.6b: Diameter profiles along axial slices for ground-truth aorta segmentations *(solid lines)* and AAA surfaces extracted by the fully sampled INRs optimized with TempReg (*dashed lines*) at available scan instances. The scan instances in days are indicated in the legends of the plots.

Interpolation and Extrapolation

In order to evaluate the INR's performance in inter- and extrapolating AAA shapes, i.e. representing AAA shapes at time points before, between and after available scan instances, we performed a series of leave-one-out experiments. In each experiment, we optimized an INR similarly as the reference experiment but left out one of the available scan instances. Using this optimized INR, we reconstruct the AAA shape at that left-out time point and compare it to the reference AAA shape.

In Figure 2.7, we compare the inter- and extrapolation performance of the INR for different temporal regularization strategies: NoTempReg (*top rows*), TempReg (*middle rows*), and ReLUTempReg (*bottom rows*). For five patients, we show the extracted inter- and extrapolated AAA shapes at the scan instances that were left out during training. The color indicates the surface distance to the corresponding reference AAA shape. The average surface distances (ASDs) are shown below each shape. For Patient 3 (*dashed boxes* in Figure 2.7), we compare

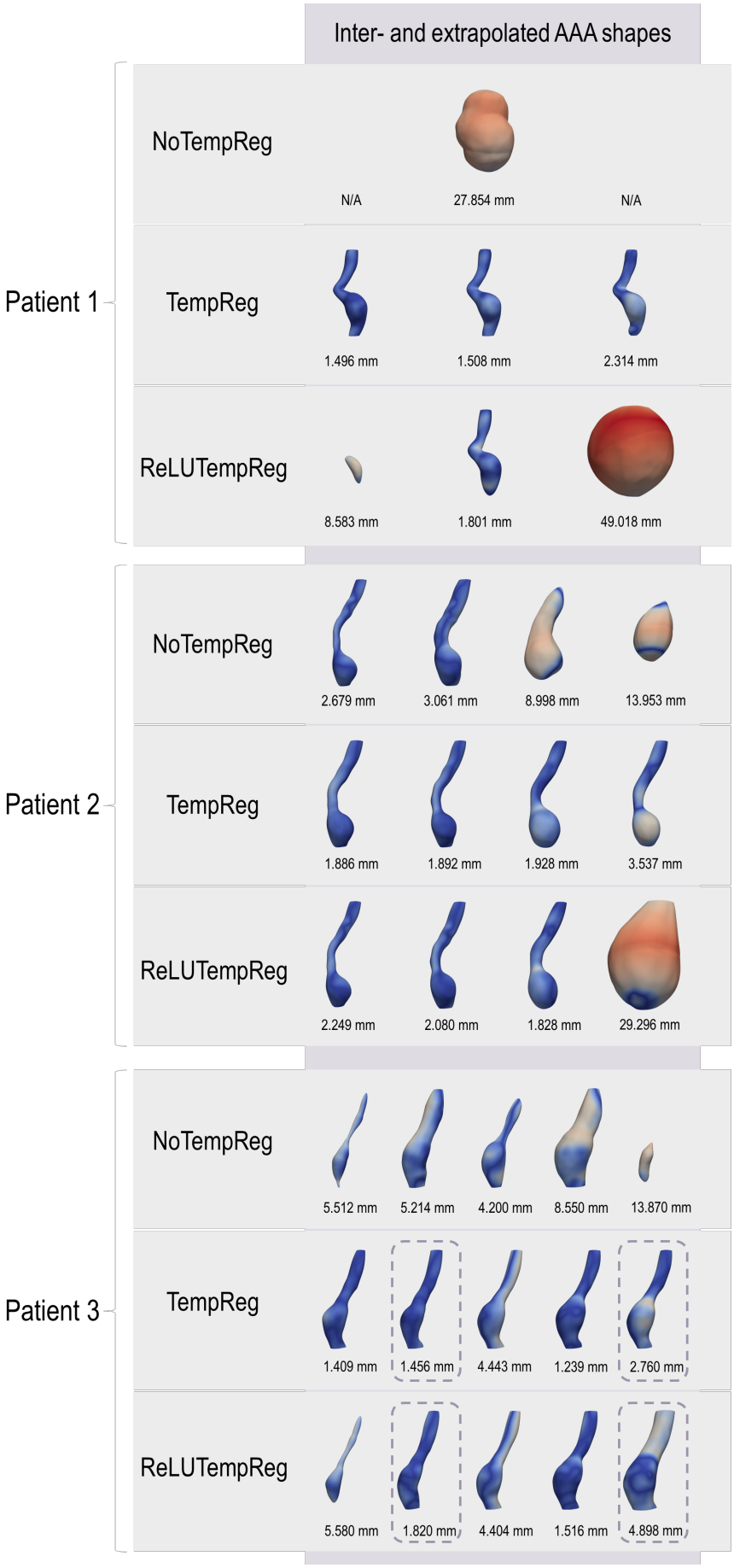
the corresponding diameter profiles along the AAA centerline of all available ground-truth aorta segmentations (*solid lines*) to inter- and extrapolated AAA shapes (*dashed lines*) as an example in Figure 2.8. In the *left*, the second scan instance $t_1 = 227$ days was left out during optimizing the INR. In the *right*, the last scan instance $t_4 = 1403$ days was left out during optimizing the INR. From Figures 2.7 and 2.8, we obtain the following key observations:

- **Temporal regularization is crucial for the INR to represent plausible AAA shapes over time.** In Figure 2.7, it can be seen that the INR model that is optimized with no temporal regularization (*top rows*) has more difficulties finding a plausible growth pattern and representing consistent AAA surfaces compared to training with temporal regularization (*middle* and *bottom rows*). Especially when the number of scan instances is small (e.g., Patient 1 and 2 having 3 and 4 scans, respectively), the model is not able to represent proper AAA shapes, thereby losing the shape of an aorta or even representing no shape at all. This is to a lesser extent for increasing number of available scan instances (e.g., Patient 4 and 5 having 6 and 7 scans, respectively). However, due to this inconsistency in representing AAA shapes when optimizing the INR with NoTempReg, we will not consider this temporal regularization strategy any further in this section.

- **An INR model that is optimized with temporal regularization can accurately *interpolate* AAA shapes evolving over time.** Both TempReg and ReLUTempReg show similar results in terms of interpolated AAA shapes and their (average) surface distance (*middle* and *bottom rows*). That is, for TempReg, the ASDs range from 0.627 to 4.443 mm, and for ReLUTempReg, the ASDs range from 0.863 to 4.404 mm.

- ***Extrapolating* AAA shapes over time is a challenging task for our INR model.** For TempReg, we observe that the model tends to reconstruct the surface of the closest known shape during training. This is shown on the *top right* in Figure 2.8 as an example. For extrapolation to the last scan instance (*dashed blue line*), the diameter profile is similar to the second last scan instance (*yellow line*). Likewise, we found that extrapolation to the first scan instance is similar to the second scan instance. As a result, the (average) surface distance is typically larger for extrapolated AAA shapes compared to interpolated AAA shapes. That is, the extrapolated ASDs range from 0.791 to 3.537 mm. For ReLUTempReg, we observe that the whole AAA shape tends to grow for extrapolated time points. For example, for Patient 3, the extrapolated diameter profile (*dashed blue line* in *bottom right* in Figure 2.8) shows that not only the dilated part of the aorta but also the healthy regions are growing over time. For Patient 1 and 2, having fewer scan instances, we observe that the extrapolated AAA shapes even grow to the end of the spatial domain (Figure 2.7). This results in substantially larger (average) surfaces distances for extrapolated AAA shapes compared to interpolated AAA shapes. That is, the extrapolated ASDs range from 2.249 to 49.018 mm.

In Figure A.2 in Appendix A, we show the diameter profiles along the AAA centerline for the ground-truth aorta segmentations (*solid lines*) and inter- and extrapolated AAA shapes (*dashed lines*) for all five patients.


Inter- and extrapolated AAA shapes
Patient 1
NoTempReg
N/A
27.854 mm
N/A
TempReg
1.496 mm
1.508 mm
2.314 mm
ReLUTempReg
8.583 mm
1.801 mm
49.018 mm
Patient 2
NoTempReg
2.679 mm
3.061 mm
8.998 mm
13.953 mm
TempReg
1.886 mm
1.892 mm
1.928 mm
3.537 mm
ReLUTempReg
2.249 mm
2.080 mm
1.828 mm
29.296 mm
Patient 3
NoTempReg
5.512 mm
5.214 mm
4.200 mm
8.550 mm
13.870 mm
TempReg
1.409 mm
1.456 mm
4.443 mm
1.239 mm
2.760 mm
ReLUTempReg
5.580 mm
1.820 mm
4.404 mm
1.516 mm
4.898 mm

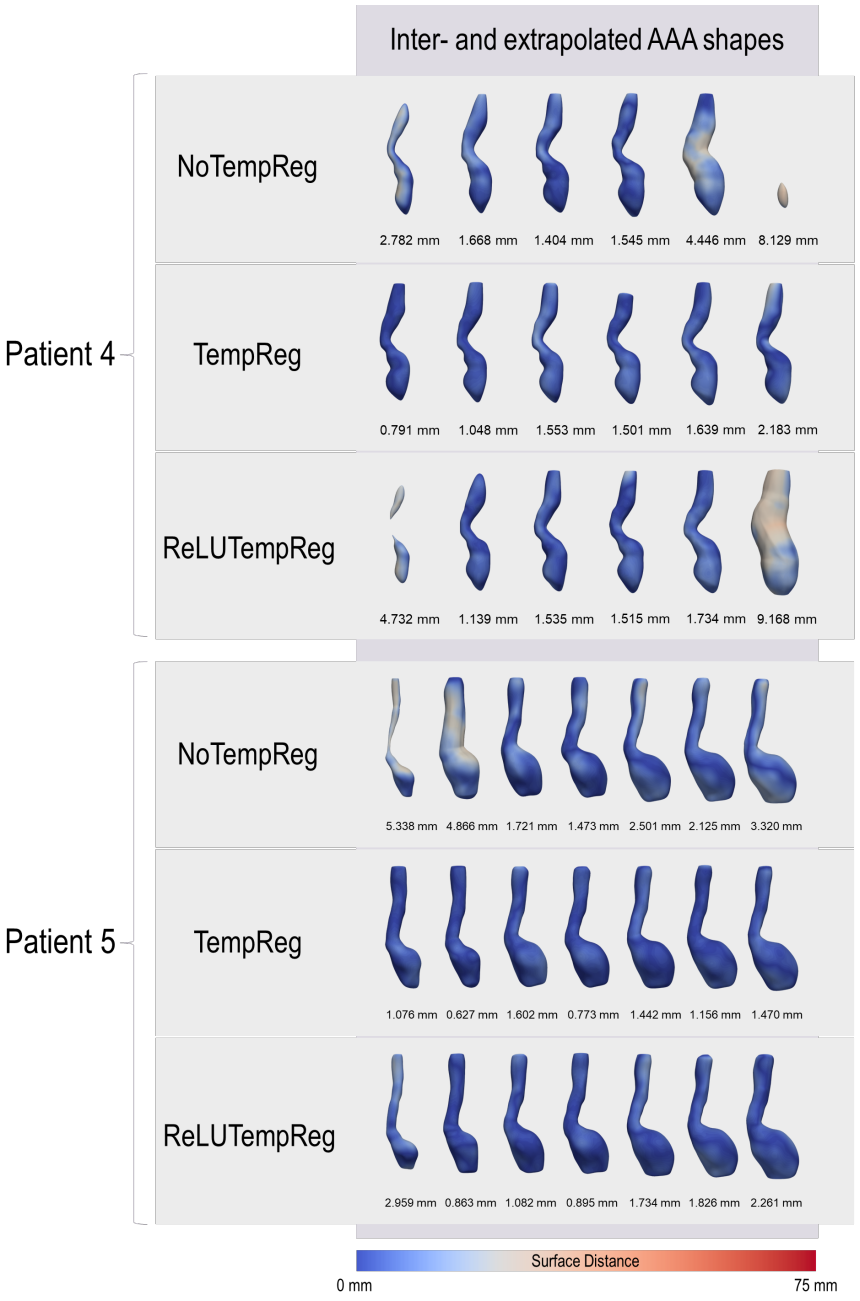


Figure 2.7: Inter- and extrapolated AAA shapes extracted at available scan instances for five patients. For each patient, we compare the results for different temporal regularization strategies: NoTempReg (*top rows*), TempReg (*middle rows*), and ReLUTempReg (*bottom rows*). The colors indicate the surface distances to the corresponding reference AAA shapes. The average surface distances (ASDs) are shown below each shape.

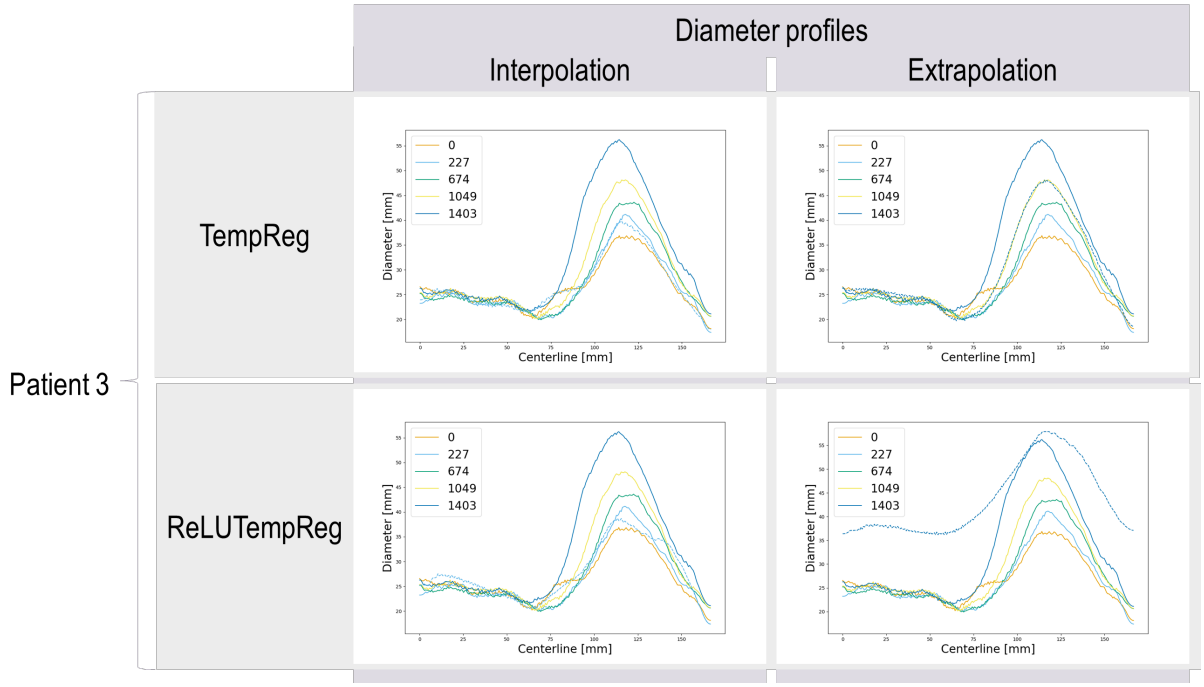


Figure 2.8: Diameter profiles along axial slices for all available ground-truth aorta segmentations (*solid lines*) and inter- and extrapolated AAA shapes (*dashed lines*) for Patient 3. We compare the results for two different temporal regularization strategies: TempReg (*top row*) and ReLUTempReg (*bottom row*). *Left:* Scan instance $t_1 = 227$ days was left out during training. *Right:* Scan instances $t_4 = 1403$ days was left out during training.

Maximum AAA Diameters

In addition to the inter- and extrapolated AAA shapes shown in Figure 2.7, we compare their maximum AAA diameters to the ground-truth maximum diameters at the available scan instances in Figure 2.9. As explained in Chapter 1, in clinical practice, assessment of rupture risk primarily relies on the maximum AAA diameter. Hence, it is important to evaluate the INR's ability to inter- and extrapolate this parameter.

For the maximum AAA diameters of the inter- and extrapolated AAA shapes, we observe the same trend as for the diameter profiles. In particular, both TempReg and ReLUTempReg can interpolate maximum diameters very well and show similar results. That is, for TempReg, the absolute difference between ground-truth and interpolated maximum diameters ranges from 0.132 to 3.870 mm, and for ReLUTempReg, this ranges from 0.111 to 2.434 mm. Note that these values are smaller than the average inter-observer variability when manually measuring the maximum AAA diameter in CT scans (4.0$\pm$5.1 mm). As observed in the diameter profiles, extrapolation is a much harder task for our INR model. For TempReg, we observe that the extrapolated maximum diameter is similar to the one of the closest known shape during training. For ReLUTempReg, we observe that extrapolation to the first scan instance results in underestimation of the maximum diameter and that extrapolation to the last scan instance results in overestimation of the maximum diameter. As a result, for both temporal regularization strategies, the absolute difference between ground-truth and extrapolated maximum diameters is much larger than for interpolated maximum diameters, namely ranging from 0.621 to 10.683 mm for TempReg, and ranging from 0.687 to 24.979 mm for ReLUTempReg.

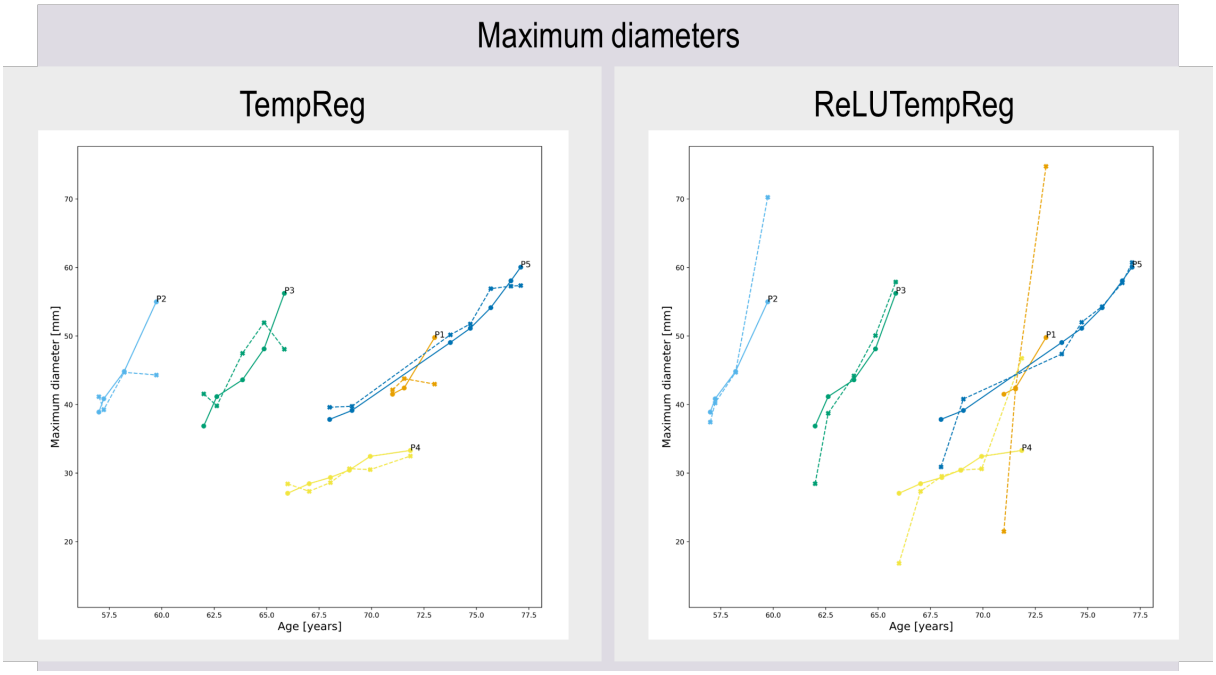


Figure 2.9: Maximum AAA diameters for ground-truth aorta segmentations (*solid lines*) and inter- and extrapolated AAA shapes (*dashed lines*) for five patients. We compare the results for two different temporal regularization strategies: TempReg (*left*) and ReLUTempReg (*right*).

2.6 Discussion

In this work, we developed a personalized approach for modeling AAA evolution, based on highly sparse and irregularly spaced longitudinal CTA data. We combine automatic methods for AAA segmentation and registration with time-dependent shape modeling using implicit neural representations (INRs). For a single patient, we represent the AAA shape evolving over time by the zero level set of its signed distance function (SDF), parameterised by space and time. We embed this time-dependent SDF in a multilayer perceptron (MLP) as an INR. In experiments with five longitudinally scanned AAA patients, we have demonstrated how this model is able to represent an evolving AAA shape at any spatial and temporal resolution. In addition, we have investigated how different temporal regularization strategies affect the inter- and extrapolation performance of the INR. With a toy problem we investigated how our model performs when the input data is free of segmentation and registration errors and where we have control over the number of time points and their intervals.

We observed that our model can accurately *interpolate* AAA shapes evolving over time. We found that temporal regularization is crucial for the INR to represent plausible AAA shapes in-between the training time points. This, however, was not the case for our toy problem; when no temporal regularization was used, the INR was able to accurately learn a growth pattern and even extrapolate on this growth pattern. Since the toy problem was based on an idealized *synthetic* dataset, this indicates that there are certain confounding factors present in the real *aneurysm* dataset that may cause the need for temporal regularization. These factors may include sparsity and irregularly spacing of our data, poor image quality, and segmentation and registration errors. Addressing these factors is expected to enhance the performance of the INR in accurately representing an AAA shape evolving over time. In particular, we found that our model is highly sensitive to errors in the initial rigid alignment of AAA shapes. That is, we observed that most AAA shapes slightly shift over time within its continuous representation obtained by the INR. In future work, we might include the renal and iliac arteries, and only register

the AAA shapes based on their healthy regions. We may also include the lumbar vertebrae which have a constant shape and size over time, as done by [16]. However, in my experience, registration based on vertebrae did not result in better alignment of the AAA shapes. This is most likely because the position of the aorta with respect to the vertebrae may vary slightly across different scans.

While our toy problem demonstrated that the model could accurately learn a growth pattern to extrapolate spheres to further time points, *extrapolation* of AAA shapes was a much more challenging task for our model. This is most likely because representing evolving AAA shapes requires the use of temporal regularization, while this was not the case for our toy problem. In particular, for the temporal regularization strategy in which we restrict any temporal change (TempReg), we observe that the extrapolated AAA surface is similar to the closest known shape during training. We hypothesise that this is due to the temporal regularization term aiming to minimize the temporal gradient of the SDF. For the temporal regularization strategy in which we allow growth but not shrinkage (ReLUTempReg), the whole AAA shape tends to grow for extrapolated time points. We hypothesise that this is due to only minimizing the positive temporal gradients of the SDF for the whole AAA shape. This allows not only the dilated part but also healthy regions of the aorta to grow for extrapolated time points, resulting in predicted AAA shapes that are not plausible. Thus, both approaches are not able to make a personalized *prediction* of AAA evolution over a future time period.

A major strength of our approach is the *continuity* of the INR. As the input coordinates in space and time are continuous, we can extract an AAA shape at any resolution and any point in time. Since we optimize a single INR for each patient, this results in highly personalized, continuous representations of AAA shapes evolving over time. These representations have potential clinical value for a more personalized assessment of AAA evolution. While longitudinal image data is now primarily used to measure the maximum diameter of the aneurysm, we are able to automatically extract the AAA shape and model its evolution over time. Furthermore, with these continuous representations, we address the problem of sparsity and irregularly spacing of longitudinal image data in predicting AAA evolution. In most longitudinal AAA studies, fixed, regularly spaced time points (e.g., a year) are obtained by linearly interpolating the AAA shapes, as done by [20]. This, however, does not take into account the patient-specific AAA growth pattern, and therefore causes a bias in predicting AAA evolution. Our model allows for interpolating AAA shapes at any moment in time according to a learned growth pattern. Moreover, since our network relies on point cloud data as input, it is independent of the imaging modality used. Therefore, longitudinal scans acquired with other imaging modalities such as MRI and 3D US can also be incorporated in this pipeline, as long as we can extract the 3D AAA shape.

A limitation of this personalized approach, however, is that it does not *generalize* easily to new AAA patients. That is, since a single INR is optimized for each patient, we cannot use this trained network for a new patient with only one CTA scan. Ideally, we would extend our model to learn regularizations from continuous representations of longitudinal AAA shapes from multiple patients. With this, we might be able to *predict* AAA evolution for new unseen patients over a desired future time period. As introduced in Chapter 1, convolutional neural networks (CNNs) have been a popular deep learning model due to their ability to learn from lots of training data. Hence, they are marked by their generalization capabilities. But since we model the evolution of AAA shapes, we are looking for an approach that operates on non-Euclidean domains. Extending CNN models to non-Euclidean domains has been an emerging research area, which is generally referred to as graph convolutional networks (GCNs). Recent work by Suk et al. has demonstrated the effectiveness of a mesh-based model that uses a GCN on artery surfaces [27]. Learning on mesh-based representations of longitudinal AAA surfaces from multiple pa-

tients, this GCN model can be optimized to predict AAA growth for new unseen patients over a desired future time period.

In conclusion, this chapter aimed to address the research sub-question: 'To what extent can implicit neural representations (INRs) be used to inter- and extrapolate AAA surfaces in time, based on sparse and irregularly spaced longitudinal CTA data?'. To answer this question, an INR model can accurately interpolate AAA shapes evolving over time when optimized with temporal regularization. Extrapolating AAA shapes, on the other hand, is a much more challenging task for an INR model. We hypothesize that this is due to the temporal regularization term aiming to minimize the temporal gradient of the SDF.

In order to address the current limitation of our INR model in extrapolation and generalization, we propose to combine the strengths of the here proposed personalized INR model with a more data-driven approach such as a GCN model. In Chapter 3, I will show how we combine these two approaches and demonstrate its performance in *predicting* AAA evolution.

3 PREDICTING AAA GROWTH USING GRAPH CONVOLUTIONAL NETWORKS

In this chapter, I propose a graph convolutional network (GCN) to estimate local growth on an AAA surface mesh, based on continuous representations of evolving AAA shapes from multiple patients. I investigate to what extent this model can be used to predict AAA growth over a future time period for new patients.

3.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) have gained immense popularity in medical image analysis tasks like classification, segmentation, reconstruction and registration. CNNs have the ability to extract multi-scale localized spatial features from input data and compose them to construct highly expressive representations, using convolution, pooling and connection layers. The key benefits of *convolution* in CNNs are sparse interactions, parameter sharing and translational equivariance. The convolution kernel only interacts with subsets of the input data and is the same everywhere in the input data. This drastically reduces the number of parameters and thereby the risk of overfitting. The use of multiple *pooling* layers increases the network's receptive field, giving access to long-range information across the input data. Another major strength of CNNs are that they can be optimized with lots of training data from multiple patients, thereby increasing its *generalization* capabilities. This allows for using a trained network to perform a certain task for new patients with data not seen during training.

Although CNNs have shown their many applications for the analysis of medical images, they can only operate on regular Euclidean data (e.g., 2D or 3D grids) (*left* in Figure 3.1). That is, the image domain on which the signal operates is fixed and thus the same for each data sample. Therefore, CNNs often require pre-processing steps like resizing the input image into the specific input size with certain pixel or voxel dimensions. Furthermore, their performance is often based on the resolution of the image data.

Another limitation of CNNs is that convolutions are not equivariant to many symmetries in the data such as rotation, reflection and scaling. Therefore, data augmentation is often necessary to obtain a network whose predictions are robust against these transformations. This, however, is inefficient because adding more training samples results in longer optimization times. Furthermore, data augmentation introduces extra hyperparameters, such as the type and extent of the transforms, which need to be carefully chosen. Instead of modifying your data to symmetries in the neural network, you would rather have a neural network that exploits symmetries in the input data.

To address these limitations, extending CNN models to non-Euclidean domains has been an emerging research area. This is generally referred to as graph convolutional networks (GCNs). In GCNs, not only the signal but also the domain on which the signal operates can be different for each data sample. That is, they operate on graphs; a type of data structure that consists of a

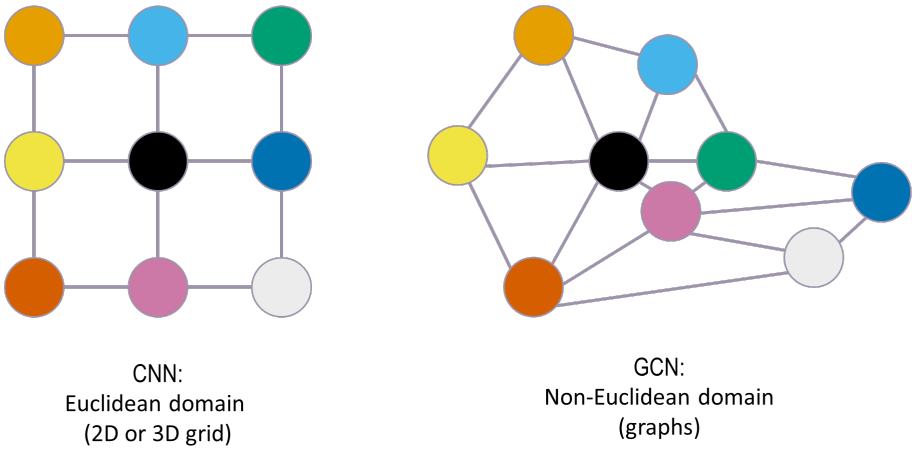


Figure 3.1: Comparison of the domains for CNNs (*left*) and GCNs (*right*). While a CNN is applied to a graph in the Euclidean domain (e.g., a 2D or 3D grid), a GCN is applied to a graph in non-Euclidean domain.

set of objects (vertices $\mathcal{V}$) and their connections (edges $\mathcal{E}$) (*right* in Figure 3.1). Note that image data structures can be regarded as instances of graphs; in a 3D grid, each voxel corresponds to a node in the graph, and the connectivity between neighboring voxels forms the edges of the graph. The general idea behind GCNs is to generalize the key benefits of convolution in CNNs to graphs.

3.2 Graph Convolutional Networks

In our work, we are interested in predicting AAA growth based on the aneurysm's shape. In Chapter 2, we obtained highly personalized, continuous representations of 3D AAA shapes evolving over time. In particular, by sampling the SDF at a regular grid and using marching cubes, we can reconstruct a mesh of the patient's AAA surface at any spatial and temporal resolution. This way of representing an AAA shape results in smooth boundaries and provides information about surface connectivity. This is not the case for voxel-based representations that are often obtained using automatic segmentation models. Moreover, since we can extract an AAA surface mesh at any time point, we are no longer restricted to the fixed, highly sparse and irregularly spaced time points at which image data was acquired. Therefore, a GCN that operates on surface meshes would be a more suitable choice for predicting AAA growth compared to regular CNNs that are limited by a fixed Euclidean domain.

3.2.1 Learning on 3D surface meshes

While CNNs take as input voxel-based images, GCNs take as input a surface mesh with descriptive input features. Let $\Omega \subset \mathbb{R}^3$ be the AAA shape and $\partial\Omega$ its 2D boundary representing the AAA's outer wall. The surface mesh $\mathcal{M}$ is a discretisation of $\partial\Omega$ that can be described by a tuple of vertices and faces $\mathcal{M} = (\mathcal{V}, \mathcal{F})$. A face is a closed set of edges $\mathcal{E}$; they consist of triangles having three edges, quads having four edges, or other simple convex n-polygons having n edges. Both the vertices and edges of a mesh can have features x_v and e_{vw}, respectively. A GCN is informed by these mesh properties and processes signals intrinsically on the mesh.

CNNs are designed for processing signals in a Euclidean domain, such as images that are defined on a 2D or 3D grid (*left* in Figure 3.1). Note that each vertex has a fixed number of neighboring vertices arranged in a regular grid where the orientation of the vertices with respect

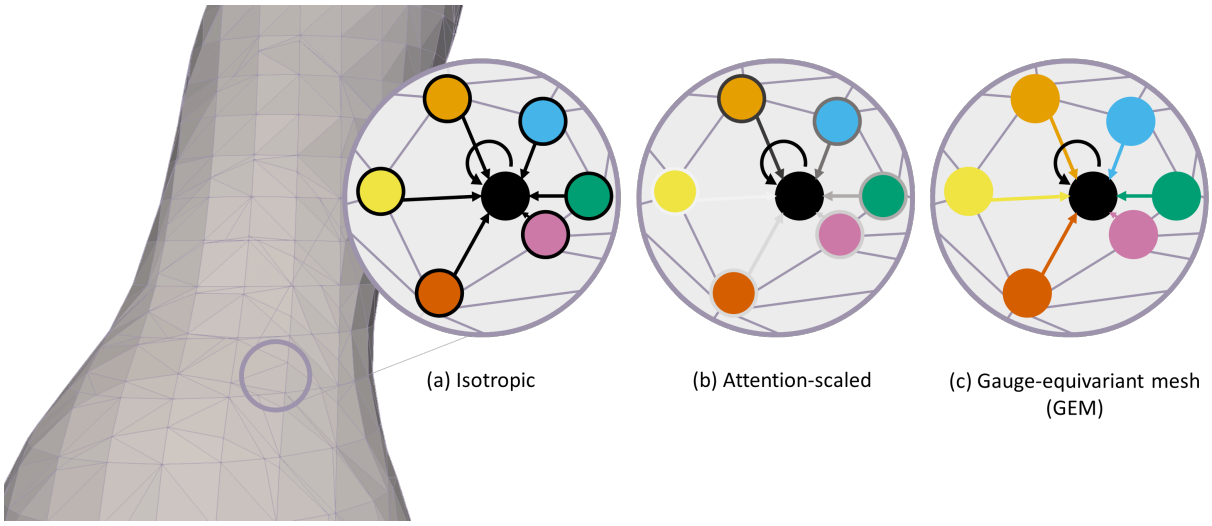


Figure 3.2: Comparison between isotropic, attention-scaled, and GEM's anisotropic convolution kernels. While isotropic convolution filters (a) process all signals mapped to the neighboring vertices in the same manner, anisotropic filters (b and c) process them distinctly. Attention-scaled convolutions (b) learn to distinguish neighboring vertices through an attention mechanism, whereas GEM convolution (c) is equipped with a notion of direction. Based on [27].

to each other is known. This is not the case for GCNs; it has no information about the orientation of the vertices with respect to each other, and the number of neighboring vertices can differ for each vertex. This poses a challenge for applying fixed-size convolutional filters. To address the lack of orientation and the variable neighborhood size in graphs, GCNs propagate information differently than CNNs. That is, graph convolutions involve the following two steps:

1. **Message passing.** Suppose $h_v^{(i)}$ represents the node embeddings for vertex v at iteration i. Note that $h_v^{(0)} = x_v$. The message $m_v^{(i+1)}$ aggregates information from the neighborhood $N(v)$ as follows:

$$m_v^{(i+1)} = \sum_{w \in N(v)} f_{message}^{(i)}(h_v^{(i)}, h_w^{(i)}, e_{vw}), \tag{3.1}$$

   where $N(v)$ is the neighborhood of vertex v.

2. **Vertex update function.** The vertex update function f_{update} creates the signal update from the messages.

$$h_v^{(i+1)} = f_{update}^{(i)}(h_v^{(i)}, m_v^{(i+1)}) \tag{3.2}$$

The trainable weights of a GCN model are in the functions $f_{message}$ and f_{update}. Specifically, the weights used for aggregating neighboring node embeddings and the weights used to update the corresponding vertex embedding are optimized during the training process.

There are different convolution filters that can be applied in the message passing step (Figure 3.2):

(a) **Isotropic convolution filters.** Isotropic convolution filters process all signals mapped to the surrounding vertices in a neighborhood in the same manner:

$$m_v^{(i+1)} = \frac{1}{|N(v)|} \sum_{w \in N(v)} h_w^{(i)} \tag{3.3}$$

(b) **Attention-scaled convolution filters.** Attention-scaled convolutional filters are anisotropic kernels that process the signals mapped to the neighboring vertices differently. They weigh each neighboring vertex with a trainable coefficient α_{vw}, resulting in the following message passing function:

$$m_v^{(i+1)} = \sum_{w \in N(v)} \alpha_{vw} h_w^{(i)} \tag{3.4}$$

(c) **Gauge-equivariant mesh (GEM) convolutional filters.** GEM convolutional filters are also anisotropic kernels. As opposed to attention-scaled convolution, GEM convolution is also equipped with a notion of direction. In the following section, we will elaborate on this type of convolution.

3.2.2 Gauge-Equivariant Mesh Convolution

We aim to estimate local growth on an AAA surface mesh. In particular, we aim to predict a 3D deformation vector for each vertex in the mesh, which represents the local growth of the AAA. Since we predict 3D vectors, we want the network to be rotation equivariant. That is, if we rotate the AAA surface mesh, we want the output to be affected in the same way. This, however, is difficult due to the lack of orientation of the neighboring vertices. To address this, we implement anisotropic kernels using gauge-equivariant mesh (GEM) convolution, as done in [34, 27].

For GEM convolution, each neighbor vertex w in a certain neighborhood $N(v)$ is projected to a tangent plane, and expressed in terms of polar coordinates with a radius and angle (r_w, θ_w) on this plane. Due to the different angles, we can distinguish between different neighboring vertices and process their signals differently. In particular, the convolution kernel sums for each neighbor $w \in N(v)$, the product of the features at w and kernel $K(\theta_w)$. The key benefits of these GEM convolutions are $SE(3)$-equivariance. That is because the tangent planes rotate with the geometric model of the AAA shape. The surface geometry is intrinsically described and does not depend on how the mesh vertices are embedded in $\mathbb{R}^3$, i.e. on its orientation in an ambient space.

3.2.3 Pooling in GCNs

An important operation in regular CNNs is pooling, the process of downsampling while preserving relevant information. The use of multiple pooling layers exponentially increases the network's receptive fields, giving access to long-range information across the input data. In GCNs, pooling layers are not widely accepted because the concept of pooling, as commonly used in CNNs, does not directly translate to graph-structured data. Pooling operations in CNNs are designed to downsample feature maps, reducing their spatial dimensions while preserving the relevant features. This is achieved by aggregating local information in a fixed-size neighborhood, which works well for grid-like data such as images. However, graphs lack a grid-like structure and have irregular connectivity patterns, making it challenging to define a meaningful neighborhood for pooling.

Instead of pooling layers, alternative strategies have been developed to access long-range information across the graph. In this work, we apply pooling in the same way as is done by Suk et al. [27]. We sample a hierarchy of vertex subsets $\mathcal{V} = \mathcal{V}_0 \subset \mathcal{V}_1 \subset ... \subset \mathcal{V}_n$, where n is the number of pooling levels. In each pooling level, the radius of the GEM convolutional filter increases ($r_0 < r_1 < ... < r_n$), giving access to a larger neighborhood. In each neighborhood, we only aggregate information from the corresponding subset of vertices. Thus, in the first pooling level, we define our GEM kernel size with radius r_1 and only aggregate information from

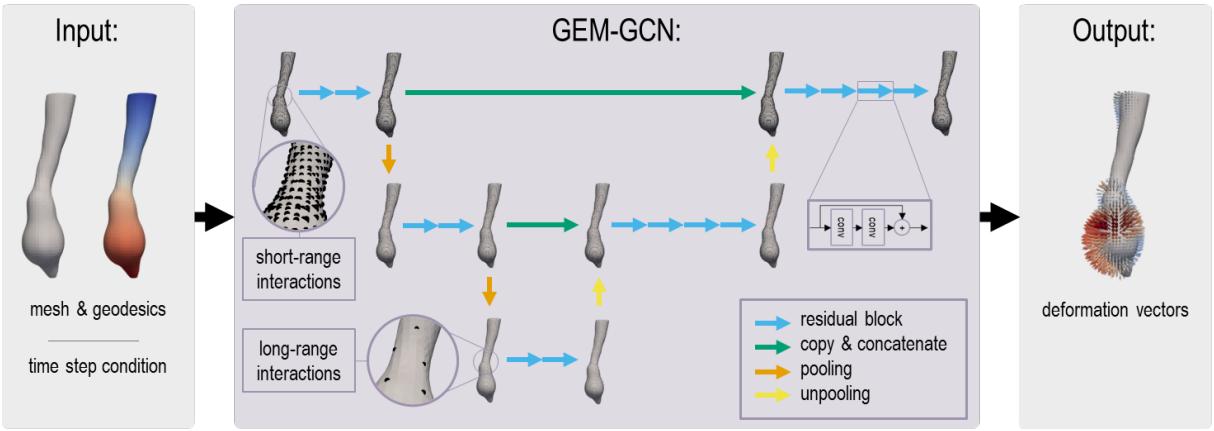


Figure 3.3: Schematic representation of our GEM-GCN, taking an AAA mesh, vertex-wise geodesic distance to the vessel inlet, and time step condition as input, outputting deformation vector fields $f^{out} : \mathcal{V} \to \mathbb{R}^3$ mapped to the vertices. Based on [27].

vertices $\mathcal{V}_1$. In the second layer, the radius r_2 becomes larger and we aggregate information from a smaller subset of vertices $\mathcal{V}_2$, and so on. Note that sampling a hierarchy of vertex subsets and defining the corresponding radii for the GEM kernels are hyperparameters that have to be chosen carefully.

3.2.4 Network Architecture

We propose a mesh-based GCN that takes as input a scalar or vector field of features mapped to vertices of the input graph $f^{in} : \mathcal{V} \to \mathbb{R}^{c_{in}}$ and outputs vector-valued predictions $f^{out} : \mathcal{V} \to \mathbb{R}^3$ mapped to the same vertices. The input graph is an AAA surface mesh, and the predicted vectors represent the magnitude and direction of AAA growth over a certain future time period. Figure 3.3 visualises the network architecture of our GEM-GCN. Like a regular CNN, our GCN consists of convolution, pooling and connection layers. In order to enable the flow of long-ranged information across the aneurysm surface, we use an encoder-decoder architecture with three pooling levels and "copy and concatenate" connections between corresponding layers in the contracting and expanding pathway. To prevent vanishing gradients, we use residual blocks consisting of two convolution layers and a skip connection. This proposed network is based on the work of [27].

3.3 Data

In this work, we use the same longitudinal *aneurysm* dataset and apply the same pre-processing as described in Section 2.4. The difference is that we now include data of all 25 AAA patients. Thus, for each patient, we have a sequence of J point clouds $\{\mathcal{X}_{t_j}\}_{j=1,\dots,J}$, where $\mathcal{X}_{t_j} = \{\boldsymbol{x}_i\}_{i \in I} \subset [-1, 1]^3$ represents points on the AAA surface of a single patient at time t_j. Based on these longitudinal point clouds, we will obtain continuous representations of AAA surfaces evolving over time using implicit neural representations (INRs). From these representations, we can extract AAA surface meshes at any spatial resolution and any point in time, which be used as input for the GCN.

3.3.1 Surface Meshes

For each patient, we optimized a single time-dependent INR based on the point clouds at all available scan instances, which was referred to as the reference experiment in Section 2.5.2.

For this, we used the temporal regularization strategy that restrict any temporal change (TempReg) as this showed the best results in Section 2.5.2. Using these fully sampled INRs, we can extract AAA surface meshes at any desired resolution and at any point in time. In Figure A.1 in Appendix A, we show diameter profiles along the AAA centerline for the ground-truth aorta segmentations (*black dashed lines*) and interpolated AAA shapes extracted at time points spaced half a year apart (*colored solid lines*) for all 25 included patients. These plots demonstrate that the fully sampled INRs can accurately represent and interpolate AAA shapes evolving over time. These INRs are therefore suitable to extract plausible AAA meshes at any time point.

In this work, we propose to predict local AAA growth over a future time period. For each patient, we have a continuous representation of the AAA shape evolving over time. From this representation, we can extract an AAA surface mesh at an initial time point t_0 and at a desired future time point t_1. Let $\Omega \subset \mathbb{R}^3$ be the AAA shape and $\partial\Omega$ its 2D boundary representing the AAA's outer wall. A surface mesh $\mathcal{M}$ is a discretisation of $\partial\Omega$ that can be fully described by a tuple of vertices and faces $\mathcal{M} = (\mathcal{V}, \mathcal{F})$. Both extracted AAA surfaces at t_0 and t_1 are described as $\mathcal{M}_{t_0}$ and $\mathcal{M}_{t_1}$, respectively. $\mathcal{M}_{t_0}$ will be the input surface mesh for our proposed GCN. $\mathcal{M}_{t_1}$ will be used to compute deformation vectors mapped to the vertices of input mesh $\mathcal{M}_{t_0}$, representing local AAA growth from t_0 to t_1. Note that we do not have any restriction in choosing the initial time point t_0 and the desired future time point t_1 due to our continuous representations of AAA shapes evolving over time. This allows us to condition our GCN on any desired time step over which we want to predict local AAA growth.

3.3.2 Deformation Vectors

For each vertex in the input surface mesh $\mathcal{M}_{t_0}$, we compute the nearest neighbor in the AAA surface mesh at a desired future time point $\mathcal{M}_{t_1}$, using the KDTree algorithm. The magnitude and direction of the resulting deformation vectors represent the local AAA growth. These deformation vectors are the labels that we aim to predict for a certain input surface mesh using our proposed GCN.

3.3.3 Input Features

We use the input surface mesh $\mathcal{M}_{t_0}$ to construct input features to the GCN. These input features with c_{in} channels describe the local shape as well as global properties.

Matrix features Like [27], we compute a surface normal for each vertex in the input surface mesh, and then construct three matrices that describe the local neighborhood. For each of the three sets of (3×3)-matrices, we take the average of the neighbrhood. In contrast to surface normals, these resulting input features define meaningful local surface descriptors that are not $SO(2)$-invariant. The vanilla surface normal would be constant in any coordinate system induced by the surface normal. Since the surface normal describes the local surface orientation in an infinitesimally small neighborhood, i.e. the precise local curvature of the aneurysm wall $\partial\Omega$, it is the preferred input feature for conventional message passing formulations. This, however, is not the case for GEM-GCNs and therefore we need the average over the neighborhood as a *curvature* descriptor.

Shortest geodesics distance to vessel inlet We extract the most cranial vertices from each input surface mesh, which we refer to as the vessel inlet of the abdominal aorta. Like [27], we append the shortest geodesic distance from each vertex to this vessel inlet as a scalar to the input features, which we compute with the vector heat method [35]. This input feature gives information about the *orientation* of each vertex on the mesh.

Time step By conditioning the GCN on time step $\Delta t = t_1 - t_0$, we can predict AAA growth over any desired time period. We append the time step Δt in days over which we want to predict AAA growth as a scalar to the input features. Thus, each vertex of a single input surface mesh has the same time step feature.

3.4 Experiments and Results

We evaluate to what extent a GCN can predict local AAA growth on an aneurysm shape using the model described in Section 3.2, which has around 10^6 trainable parameters. The 25 included patients were randomly split 18:5:2 into training, validation, and test sets, respectively. The network was trained using an L^1 loss and Adam optimizer with batches of 18 input surface meshes and a constant learning rate of 10^{-3}. We use ReLU activation functions and employ batch normalisation before each activation. The radii for the GEM kernels in the three pooling layers were set to $r_0 = 3.33$, $r_1 = 6.66$, and $r_2 = 9.99$ mm, and we sampled a hierarchy of vertex subsets using the ratios 1, 0.4, and 0.1.

In each epoch, we sampled a single batch containing one input surface mesh with corresponding input features for each of the 18 training patients. This sampling involved the following steps (Figure 3.4):

1. We randomly sampled a time step Δt between 182.5 and 730 days (corresponding to 0.5 and 2 years, respectively). For patients with a total follow-up time smaller than 2 years, we randomly sampled a time step Δt between 182.5 days and the total follow-up time. Based on the sampled time step, we randomly sampled an initial time point t_0 between 0 days and the total follow-up time of the particular patient minus the sampled time step.

2. At time point t_0, the input surface mesh $\mathcal{M}_{t_0}$ is extracted from the fully sampled INR for the particular patient. We also extracted the surface mesh $\mathcal{M}_{t_1}$, where $t_1 = t_0 + \Delta t$.

3. We then obtained the deformation vectors by computing the nearest neighbor for each vertex in the input surface mesh $\mathcal{M}_{t_0}$ to $\mathcal{M}_{t_1}$. The resulting deformation vectors are the labels. The input features, namely matrix features and geodesics, were computed as described in Section 3.3.3. Together with the AAA surface mesh $\mathcal{M}_{t_0}$ and time condition Δt, they form the input to the GCN.

This way of randomly sampling time steps Δt and input surface meshes $\mathcal{M}_{t_0}$ during training ensures that the network sees new data in each epoch. This enhances the network's generalization and robustness, leading to improved performance on unseen validation and test data. The network was trained for 400 epochs on NVIDIA A40 GPUs and parallelization over two GPUs was necessary to fit the batches into memory. While training took around 1:10 [h], inference for a previously unseen AAA mesh takes less than 2 s including pre-processing.

In this section, we will show the results for the five validation patients, namely P8, P23, P25, P22, and P17, with their number of CTA scans ranging from 4 to 7 (Figure 2.3). For clarity, we will refer to them as Patient 1, Patient 2, Patient 3, Patient 4, and Patient 5, respectively. We show the results for these validation patients and not the two test patient because both test patients have only two scan instances and therefore allow quantitative evaluation for only one scan instance per patient. Our validation patients, having more scan instances, provide a more comprehensive evaluation of the GCN's performance.

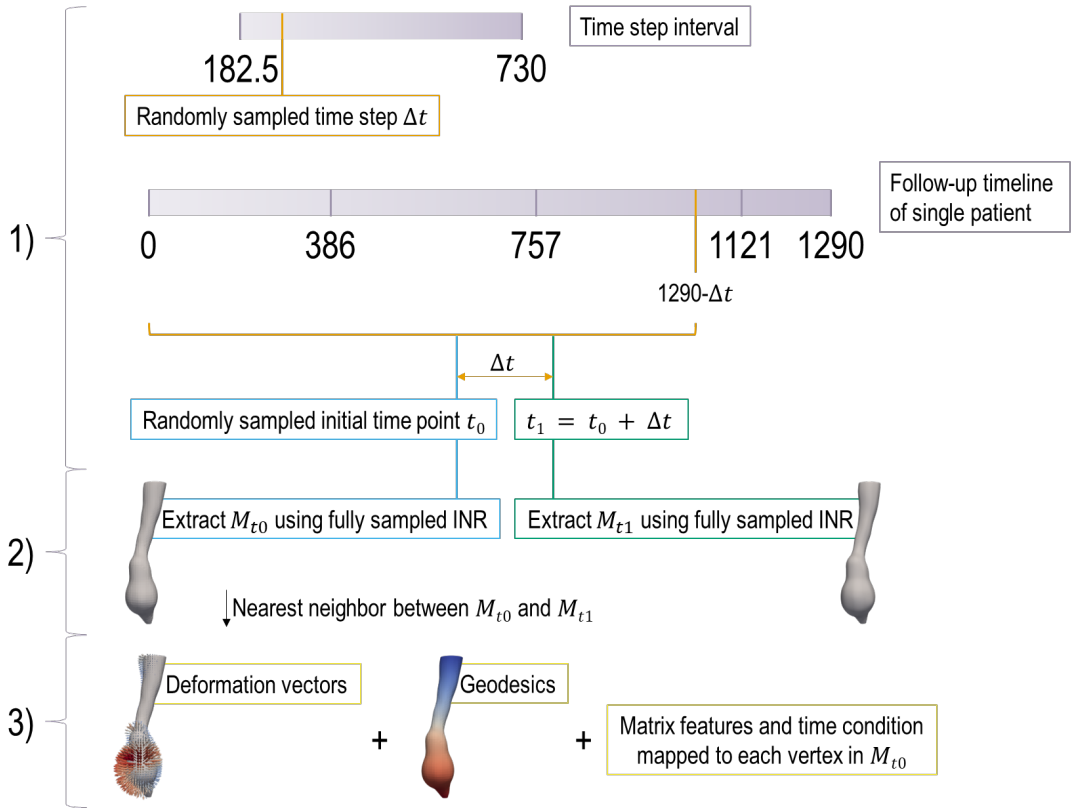


Figure 3.4: Sampling of an input surface mesh with corresponding input features involves the following steps: 1) sampling of of time step Δt, initial time point t_0 and next time point t_1, 2) extracting AAA surface meshes $\mathcal{M}_{t_0}$ and $\mathcal{M}_{t_1}$, and 3) computing and mapping corresponding labels (deformation vectors) and input features (geodesics, matrix features, and time condition) to the vertices of $\mathcal{M}_{t_0}$.

3.4.1 Time Conditioning

In Figure 3.5, we compare the deformation vectors predicted by the trained GCN model for different time step (ranging from 91.25 to 1095 days) for the five validation patients. We observe that a larger time step condition results in larger deformation vector predictions; note that the magnitude of the deformation vectors increase for increasing time steps (from *left* to *right* in Figure 3.5). Furthermore, the model demonstrates its ability in predicting AAA growth specifically in the dilated part of the aorta, leaving the healthy parts unaffected.

We observe that the model is able to extrapolate to time step conditions outside the interval during training (indicated in *red* in Figure 3.5). For example, for Patient 1, 3, and 4, it predicts larger deformation vectors for $t = 1095$ days than for $t = 730$ days, even though the model was trained with time steps between 182.5 and 730 days. This means that the model is able to learn a growth pattern depending on the time condition. This, however, is not the case for Patient 2 and 5, for which the deformation vector for $t = 1095$ days stayed the same or even decreased compared to $t = 730$ days.

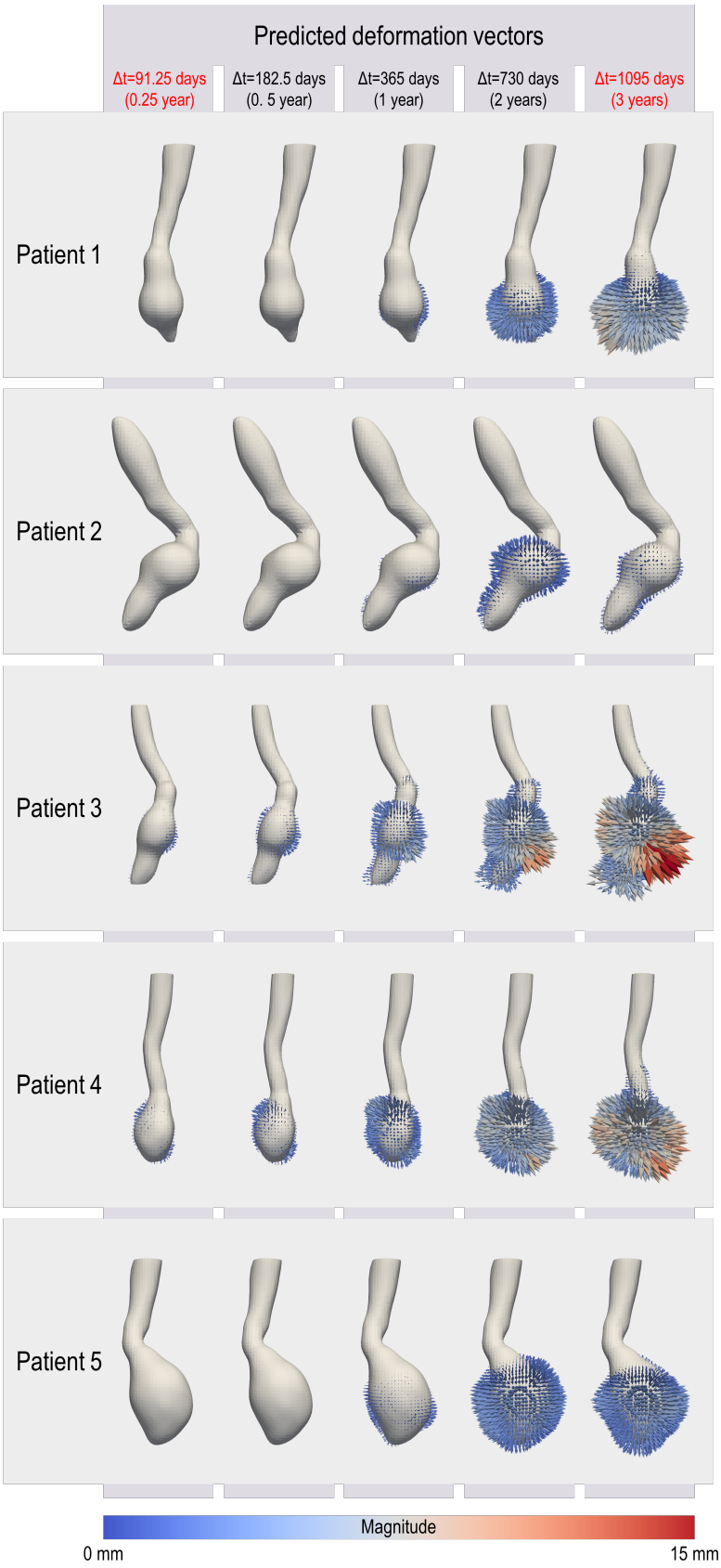


Figure 3.5: Predicted deformation vectors on AAA shapes of the five validation patients for increasing time steps from *left* to *right* (91.25 to 1095 days). Red time steps indicate extrapolated time steps that are outside the interval during training (182.5-730 days).

3.4.2 Quantitative Evaluation

In order to quantitatively evaluate the GCN's performance in predicting AAA growh, we need to compare the predictions with ground-truth data. Note that the most trustworthy ground-truth data are the CTA data at fixed scan instances. This means that we have to predict AAA growth from one scan instance to a later scan instance, and compare the predicted AAA shape with the ground-truth AAA shape at that later scan instance. We create the predicted AAA shapes by adding the predicted deformation vectors to the vertices in the input surface mesh. Since the fully sampled INRs perfectly represented AAA surfaces at the scan instances (Figure 2.6b), we use these reference AAA shapes to evaluate the predicted AAA shapes. As in Chapter 2, we use (average) surface distances, diameter profiles, and maximum AAA diameters to evaluate the prediction performance of our GCN model.

Predicted AAA Shapes

In this experiment, we predicted the AAA shapes from the reference AAA shapes at the previous scan instance. In particular, from the reference AAA shape at scan instance t_0, we predict the AAA shape at the following scan instance t_1. And from the reference AAA shape at this scan instance t_1, we predict the AAA shape at the following scan instance t_2, and so on. Note that the time step conditions are the differences between the scan instances (e.g., $\Delta t_2 = t_2 - t_1$). We visualize this in Figure 3.6, where we show the predicted AAA shapes at the scan instances for the five validation patients. The color indicates the surface distance to the corresponding reference AAA shape extracted from the fully sampled INR at the same time point. The average surface distances (ASDs) are shown above each shape. The reference AAA shapes are overlaid transparently. The time points t and time steps Δt in days are indicated below the AAA shapes; red indicates extrapolated time steps that are outside the interval during training (182.5-730 days).

We observe that, for some patients, the reference and predicted AAA shapes were not perfectly aligned. For example, for Patient 3, the predicted AAA shape at $t_2 = 801$ days is shifted to the left compared to the reference AAA shape overlaid transparently. We hypothesize that this misalignment may be caused by registration errors during pre-processing. This misalignment may result in large ASDs; for Patient 3, the predicted AAA shape at $t_2 = 801$ days has the largest ASD value of 4.083 mm. The ASDs may therefore not accurately reflect the performance of the GCN model in predicting local AAA growth. To properly evaluate the aneurysm growth, regardless of slight translations, we compare the diameter profiles.

Diameter Profiles

For Patient 1 and 3 (*dashed boxes* in Figure 3.6), we compare the diameter profiles along the AAA centerline of all available ground-truth aorta segmentations (*solid lines*) to the AAA surfaces predicted by the GCN model (*dashed lines*) as an example in Figure 3.7 In each plot, we compare three diameter profiles: the ground-truth and predicted diameter profiles at t_j, and the ground-truth diameter profile at t_{j-1} from which the prediction was made. For Patient 3, which had the largest ASD value for the predicted AAA shape at $t_2 = 801$ days (Figure 3.6), we observe that the corresponding diameter profile has increased with respect to the previous scan instance $t_1 = 258$ days (*right column*, *second plot* in Figure 3.7). The predicted dashed line approximately aligns with the ground-truth solid line, thereby reflecting a good performance of the GCN model in predicting AAA growth. The diameter profiles are therefore a better way to evaluate the GCN's performance in predicting local AAA growth, compared to the ASDs.

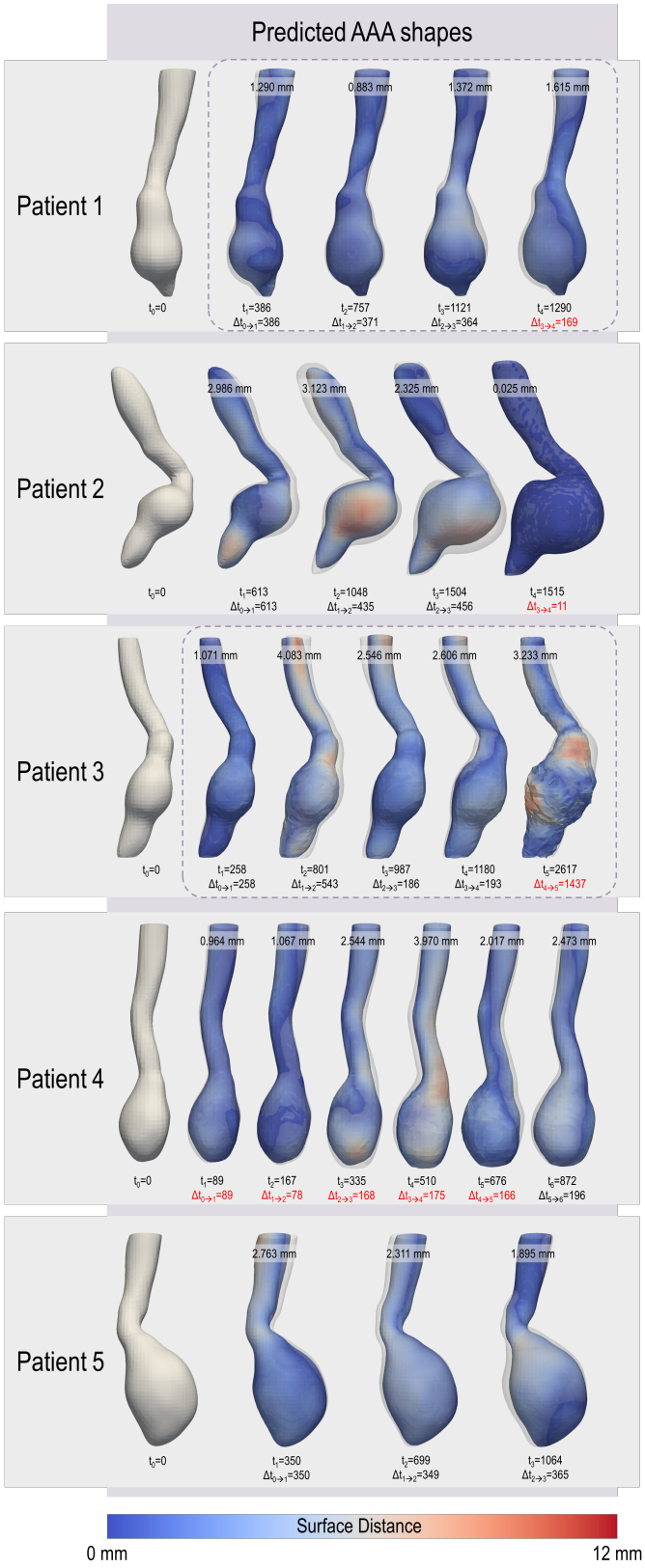


Figure 3.6: Predicted AAA shapes at scan instances for the five validation patients. The colors indicate the surface distances to the corresponding reference AAA shapes extracted from the fully sampled INRs at the same time point. The average surface distances (ASDs) are shown above each shape. The reference AAA shapes are overlaid transparently. The time points t and time steps Δt in days are indicated below the AAA shapes; red indicates extrapolated time steps that are outside the interval during training (182.5-730 days). Note that at time point $t = 0$, only the reference AAA shape is shown.

From the diameter profiles, we obtain the following key observations:

- **While our GCN model is able to predict AAA growth in the right direction, it often tends to underestimate the growth.** In the dilated part of the aorta, the predicted diameter profiles are often larger than the previous ground-truth diameter profile. The diameters, however, are often underestimated, predicting too little AAA growth compared to the ground-truth. For example, for Patient 1 (*left column*), each predicted diameter profile at t_j lies above the previous ground-truth diameter profile at t_{j-1}, but under its ground-truth diameter profile at t_j.
- **Our GCN model specifically predicts AAA growth in the dilated part of the aorta, leaving the healthy parts unaffected.** In particular, whereas the predicted diameters at the dilated part are often larger than the previous ground-truth diameter profile, the predicted diameters at the healthy part of the aorta are approximately the same. As seen for Patient 1 and 3, all diameter profiles at the healthy part of the aorta perfectly align, and the GCN model only predicts growth in the dilated part. This, however, was not reflected by the ASDs in Figure 3.6. For example, for Patient 3, the predicted AAA shapes at $t_3 = 987$ and $t_4 = 1180$ days have large surface distances at the top (healthy) part of the aneurysm, resulting in large ASDs. This also indicates that ASDs do not accurately reflect the performance of the GCN model in predicting local AAA growth. These errors in the surface distances are most likely caused by registration errors.

In Figure A.3 in Appendix A, we show the diameter profiles along the AAA centerline for the ground-truth aorta segmentations (*solid lines*) and predicted AAA shapes (*dashed lines*) for all five validation patients.

Maximum AAA Diameters

In addition to the predicted AAA shapes shown in Figure 3.6, we compare their maximum AAA diameters to the ground-truth maximum diameters at the available scan instances in Figure 3.8. As explained in Chapter 1, in clinical practice, assessment of rupture risk primarily relies on the maximum AAA diameter. Hence, it is important to evaluate the GCN's ability to predict this parameter.

For Patient 3 and 5, the predicted maximum diameters are very close to the ground-truth maximum diameters, except for the last scan instance for Patient 3. This is most likely due to a large time step between the second-last and last scan instance. For Patient 1, the predicted maximum diameters are larger than the previous ground-truth maximum diameters, but always smaller than its ground-truth maximum diameter at the same scan instance. This indicates underestimation of the GCN model in predicting local AAA growth. The same is observed for Patient 2, but here the underestimation is even larger. Interesting behavior is observed for predicting the AAA shape at t_1 from the first scan instance t_0. That is, for four of the five validation patients (Patient 1, 2, 3, and 5), the predicted maximum diameter at t_1 is smaller than the ground-truth maximum diameter at t_0, indicating shrinkage of the AAA shape. For almost all other scan instances, an increase in maximum diameter is observed. Overall, the absolute difference between ground-truth and predicted maximum diameters at the same time point ranges from 0.012 to 8.473 mm.

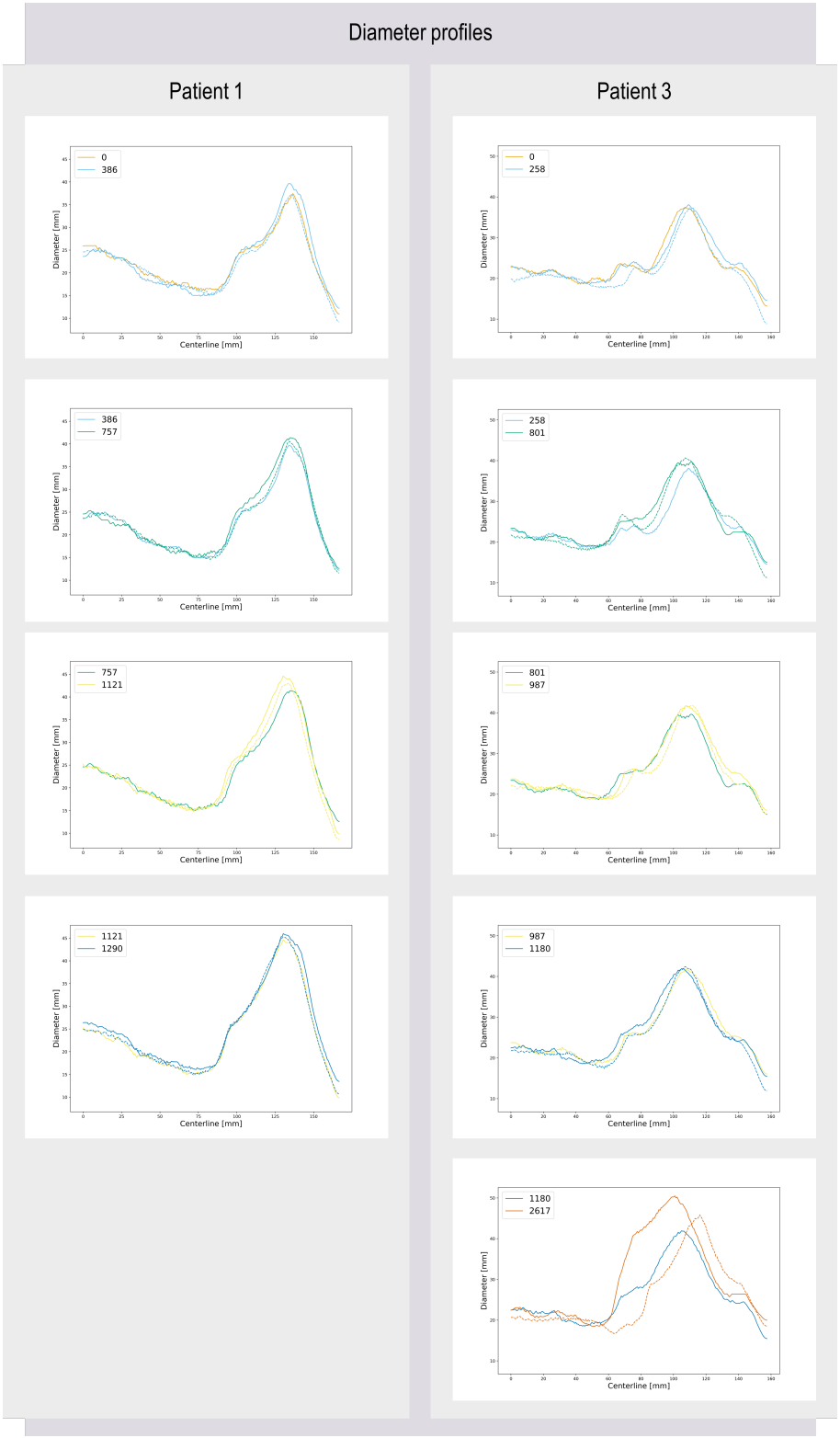


Figure 3.7: Diameter profiles along the axial slices for all available ground aorta segmentations (*solid lines*) and predicted AAA shapes (*dashed lines*) for Patient 1 and 3 (*left* and *right*, respectively). In each plot, we compare three diameter profiles: the ground-truth and predicted diameter profiles at t_j, and the ground-truth diameter profile at t_{j-1} from which the prediction was made.

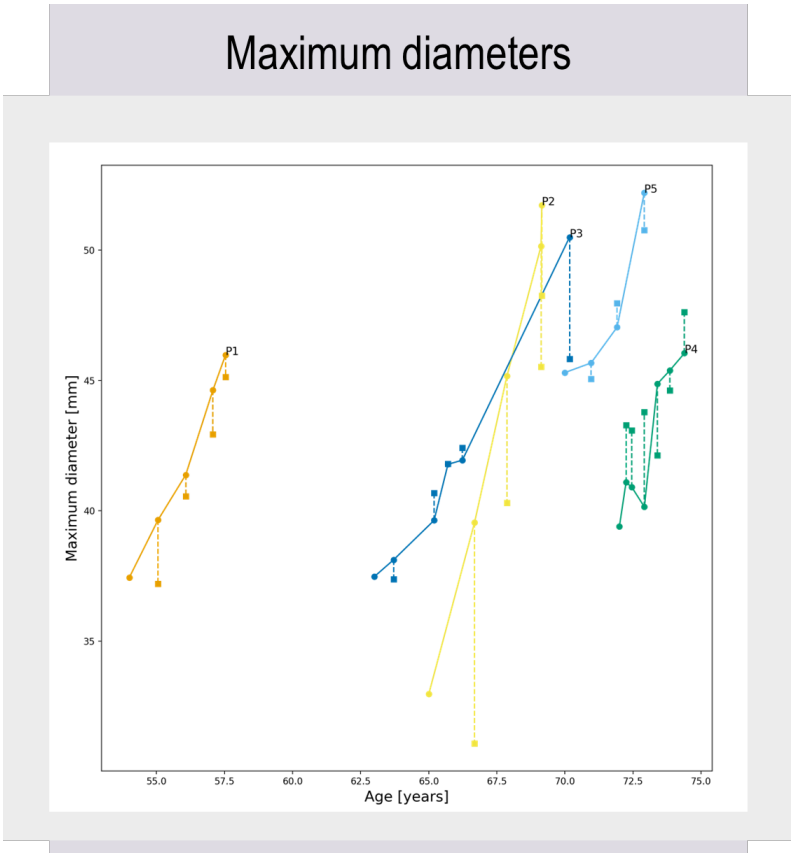


Figure 3.8: Maximum AAA diameters for ground-truth aorta segmentations (*solid lines*) and predicted AAA shapes (*squares*) for the five validation patients. The *dashed lines* indicate the difference between ground-truth and predicted maximum diameters at the same time point.

3.5 Discussion

In this work, we have presented an $SE(3)$-equivariant graph convolutional network (GCN) for the prediction of local AAA growth, operating on surface mesh representations of the AAA's outer wall. We use continuous representations of evolving AAA shapes from multiple patients, that we obtained using implicit neural representations (INRs) (Chapter 2), to train our GCN model to predict local growth on an AAA surface mesh. Since we condition our GCN model on a time step, we can predict local AAA growth over any desired future time period. For five validation patients, we have demonstrated how this model is able to predict AAA growth from one scan instance to a later scan instance.

We observed that our GCN model predicts AAA growth in the right direction specifically in the dilated part of the aorta, leaving the healthy parts unaffected. However, the model tends to underestimate AAA growth. There may be several factors that attribute to this underestimation. We hypothesize that the GCN model is underfitted by the limitid amount of longitudinal image data from multiple patients. We trained our model for 400 epochs on two NVIDIA A40 GPUs, which took around 1:10 [h]. This training time is much smaller compared to the work of Suk et al. [27] on which our model is based. They had a training time until convergence of 22:24 [h] for parallelized training on the same two GPUs. But if we train for much more epochs ($\sim$2000 epochs), we observed that the GCN's performance in predicting local AAA growth decreased for the validation patients, indicating overfitting. Since our aneurysm dataset consists of only 25 AAA patients, of which 18 patients are used for GCN model training, the diversity of AAA shapes is limited. AAA shapes are unique, and therefore a large training dataset is required for

the GCN model to learn regularizations from these shapes, thereby capturing the wide range of variations in AAA shapes and their trends in evolution over time. Ideally, we would optimize the network until convergence based on continuous representations of evolving AAA shapes from hundreds or even thousands of patients. Another explanation for underestimation of AAA growth might be that we did not finetune the network architecture, hyperparameters, or training process enough. Examples are number of pooling layers, different kernel radii and ratios for sampling a hierarchy of vertex subsets, and the use of different losses and activation functions. In future work, we may perform an ablation study in which we analyze the effect of these individual factors on the GCN's performance in predicting local AAA growth.

A major strength of our approach is the conditioning of our GCN model on a time step. Whereas most state-of-the-art prediction models predict AAA growth over a certain fixed time period (e.g., a year), we are able to predict AAA growth over any desired future time period. With this, we are able to predict an AAA shape from one scan instance to a later scan instance, which allows for adequate quantitative evaluation by comparing to ground-truth data. Note that ground truth data for models that predict over a fixed time period are often obtained by linearly interpolating or extrapolating the AAA shape, and are therefore biased from highly sparse and irregularly spaced longitudinal image data.

Although time conditioning allows any time step as input, we did not yet evaluate how different time step intervals during training affect AAA growth predictions over certain time periods. In this work, we trained our GCN model with time steps between 182.5 and 730 days (corresponding to 0.5 and 2 years, respectively). We observed that our model is able to extrapolate to time step conditions outside the interval during training; the model predicts larger deformation vectors for time steps of 1095 days (3 years) than for smaller time steps (Section 3.4.1). This, however, was not the case for all validation patients, indicating that the model is not able to fully learn from the time conditions and extrapolate on it. In future work, we may investigate how dividing large time steps into multiple smaller steps affects the GCN's performance in predicting local AAA growth. For example, instead of predicting AAA growth over a time period of 3 years at once, we can predict two times over a time period of 1.5 years, or three times over a time period of 1 year. This also allows us to investigate to what extent errors propagate through our pipeline. Since our pipeline consists of many steps, multiple errors can accumulate throughout the pipeline and affect the GCN's performance in predicting local AAA growth. If we predict multiple times over smaller time steps, we can investigate to what extent errors accumulate in our pipeline.

In conclusion, this chapter aimed to address the research sub-question: 'To what extent can a graph convolutional network (GCN) predict local AAA growth, based on longitudinal aneurysm representations from multiple patients?'. To answer this question, a GCN model can predict AAA growth specifically in the dilated part of the aorta, leaving the healthy parts unaffected. While our GCN model is able to predict AAA growth in the right direction, it often tends to underestimate the growth. We hypothesize that this is due to the limited amount of longitudinal image data and therefore the limited diversity of AAA shapes in our dataset.

4 DISCUSSION AND CONCLUSION

In this work, I have presented novel deep learning methods for modeling and prediction of local AAA growth. Our proposed pipeline consists of automatic aorta segmentation using an nnU-Net, continuous surface representation using an implicit neural representation (INR), and predicting local AAA growth on the surface using a graph convolutional network (GCN) (Figure 1.1 in Chapter 1). We have optimized this pipeline based on highly sparse and irregularly spaced longitudinal CTA data of AAA patients. The pipeline holds significant clinical value for a more personalized assessment of AAA growth for new AAA patients. Once a new patient is diagnosed with AAA based on 3D medical image data, we can automatically segment the AAA shape based on a trained nnU-Net. From this AAA shape, we can automatically extract the maximum AAA diameter that is currently used in clinical practice to assess rupture risk. This standardized method obviates the need for manual measurements of the maximum AAA diameter in medical image data, and hence inter-observer variability associated with this parameter is no longer a point of concern. Furthermore, since we segment the whole AAA shape, we can also automatically extract other relevant geometrical parameters like AAA volume, ILT thickness and aortic tortuosity that have been associated with AAA growth [13]. If longitudinal image data is available, we can even obtain a continuous representation of the AAA shape evolving over time using an INR that is conditioned on a time coordinate. This enhances *retrospectively* assessing AAA evolution by accurate interpolation of the AAA shape at any moment in time. Moreover, our proposed pipeline exceeds retrospective assessment; that is, we are able to *predict* local AAA growth. In particular, operating on a surface mesh representation of the AAA shape, our GCN model can predict local AAA growth over any desired future time period. In this chapter, I will discuss the strengths and limitations of our proposed automatic pipeline and their potential to aid physicians in clinical decision-making. I will also elaborate on future work and the next steps that should be taken to adopt our pipeline in clinical practice.

4.1 Strengths and Limitations

A major strength of our automatic pipeline is the conditioning of the both the INR and GCN model on a *time coordinate*. In most studies, data is collected retrospectively and hence dense and regularly spaced time steps cannot be obtained. Fixed, regularly spaced time points (e.g., a year) are often obtained by linearly interpolating the AAA shapes. This, however, does not take into account the patient-specific, local AAA growth pattern, and therefore causes a bias in predicting AAA evolution. Due to the conditioning of our INR model on a time coordinate, we are able to obtain a personalized, continuous model of a 3D AAA shape evolving over time. This allows for interpolation at any moment in time according to a learned growth pattern. Furthermore, whereas most state-of-the-art prediction models predict AAA growth over a certain fixed time period (e.g., a year), we are able to predict local AAA growth over any desired future time period. With this, we can predict the AAA shape from one scan instance to a later scan instance, which allows for adequate quantitative evaluation by comparing to ground-truth data. With our time conditioning, we are also able to compare our results to other studies that predict AAA growth over different fixed time steps.

Furthermore, our pipeline makes optimal use of the strengths of two deep learning models. Due to the *continuity* of INRs, we can obtain highly personalized, continuous models of 3D AAA shapes evolving over time. But since a single INR is optimized for a single patient, this personalized approach does not generalize easily to new AAA patients. To address this, we use a GCN model that is marked for its *generalization* performance. This network is able to learn local AAA growth patterns from continuous representations of evolving AAA shapes from multiple patients. Hence, an optimized GCN model can be used to predict local AAA growth for new unseen patients. Thus, we combine highly personalized models of evolving 3D AAA shapes with a GCN model that can learn regularizations from these representations.

However, since our pipeline consists of many steps (Figure 1.1 in Chapter 1), multiple errors can accumulate throughout the pipeline. Examples are poor image quality and segmentation errors, resulting in inadequate representations of the AAA shape, registration errors, and errors made by the INR and GCN models. We found that our INR model is highly sensitive to errors in the initial rigid alignment of AAA shapes. In particular, we observed that most AAA shapes slightly shift over time within its continuous representation obtained by the INR. As a result, not only growth of the dilated part of the aorta, but also displacement of healthy regions due to registration errors are passed on as labels to the GCN model. The GCN model does not distinguish between these displacement vectors due to AAA growth and registration errors, and will learn to predict both. Fortunately, we found that our GCN model specifically predicts AAA growth in the dilated part of the aorta, leaving the healthy regions unaffected. We hypothesize that small registration errors are cancelled out during training of the GCN model because their direction is arbitrary over the entire dataset, in contrast to growth of the dilated part of the aorta. In spite of that, perfectly aligning AAA shapes using rigid registration remains a challenging task and should be considered carefully. In future work, we may include additional anatomical landmarks to enhance registration of AAA shapes, e.g., the renal and iliac arteries, and the lumbar vertebrae that have a constant shape and size over time.

Another limitation of our study is the fairly limited amount of longitudinal image data. Note that while our INR model is limited in terms of the number of available scan instances per patient, our GCN model is limited regarding the number of patients in our aneurysm dataset. This is because a single INR network is optimized based on longitudinal image data from a single patient, whereas a GCN model is optimized based on continuous representations of evolving AAA shapes from multiple patients. Since our aneurysm dataset consists of only 25 AAA patients, of which 18 patients are used for GCN model training, the diversity of AAA shapes is limited. AAA shapes are unique, and therefore a large training dataset is required for the GCN model to learn requlatizations from these shapes, thereby capturing the wide range of variations in AAA shapes and their trends in evolution over time. Ideally, we would optimize the network based on longitudinal image data from hundreds or even thousands of AAA patients. We hypothesize that this would significantly enhance the robustness, generalization capabilities, and prediction performance of our GCN model.

4.2 Clinical Impact

Accurately predicting AAA growth may aid in clinical decision-making regarding the need for surveillance and surgical intervention. Currently, AAA growth is only assessed retrospectively at fixed time points by manual measurement of maximum AAA diameter in medical image data. A model that accurately predicts AAA growth provides valuable insights into the future, allowing for more pro-active AAA surveillance such as adapting its interval. If, according to our prediction model, an aneurysm is expected to grow rapidly, surveillance intervals could be shortened

to keep a closer eye on the patient, or even surgical repair can be recommended. In case of a predicted stable aneurysm, the surveillance interval could be increased. Consequently, we can minimize unnecessary radiation exposure for the patient and reduce the burden on healthcare by less surveillance. A prospective view on patient follow-up is also beneficial for patient well-being. Instead of only looking at the current situation, a glimpse into the future may put the patient more at ease. It is important to note that our proposed automatic pipeline is meant to complement the current pipeline, rather than replace it. Our pipeline has the potential to assist clinicians by providing standardized assessment and prediction of AAA growth. Clinicians have extensive medical knowledge, experience and expertise in AAA management, and should consider the broader context of the patient when making clinical decisions.

A 3D model of an AAA shape may also aid in treatment planning. While EVAR has become a commonly used surgical intervention for AAAs, it may not be the ideal option for patients with challenging anatomical features. Anatomical characteristics that are used to determine suitability of EVAR include aortic neck length, neck diameter, suprarenal and infrarenal neck angulation, distal fixation site length, and distal fixation site diameter [36]. In clinical practice, these anatomical characteristics are measured manually in a pre-operative CT scan, and can therefore be associated with inter-observer variability. Using the segmentation model from our pipeline, we can automatically segment the 3D AAA shape from pre-operative CT scans. From this aorta segmentation, we can automatically extract anatomical characteristics that can assist in choosing the suitable type and size of different components of the stent graft, including an aortic bifurcated main body and one or two iliac limbs. Note that we should incorporate additional anatomical regions, such as renal and iliac arteries, in our pipeline to adequately assess anatomical suitability of EVAR. In addition to aiding in treatment planning, a 3D model of the AAA shape may also assist surgeons during EVAR. In particular, during X-ray guided EVAR, the aorta segmentation along with the planned stent graft components can be projected onto the X-ray images. This projection may facilitate surgeons in accurate placement of the stent graft, especially during complicated EVAR procedures like fenestrated and branced EVAR (fEVAR and bEVAR, respectively), which involve the placement of additional branches into the renal arteries, celiac trunk, or superior mesenteric artery. With this personalized treatment planning and assistance during EVAR, we may decrease long term aneurysm-related mortality, re-intervention and rupture rates after EVAR.

From a technical perspective, our pipeline has the potential to make a significant impact in various other clinical settings. There are other applications in which modeling or predicting growth is important in clinical decision-making. For example, our pipeline can be adapted to model and predict the evolution of tumor shapes over time. AAA and tumor management share similarities as both involve surveillance of patients, acquiring medical image data during follow-ups to assess size. Thus, in both cases we have sparse and irregularly spaced longitudinal image data. As long as we can extract the shape's surface at multiple moments in time, we can obtain a continuous representation of its evolution using our time-dependent INR. Using continuous representations of evolving tumor shapes from multiple patients, we can learn their growth patterns and predict future progression for new tumor shapes using our GCN model. Note that we can easily condition our GCN model to tumor-specific factors that play a crucial role its evolution by appending these factors to the input features. With a tumor prediction model, clinicians can optimize surveillance and treatment planning, assess the response to treatments, and estimate prognosis.

4.3 Future Work

4.3.1 ILT Thickness

In future work, we could improve the GCN's performance in predicting local AAA growth by including more relevant input features such as intraluminal thrombus (ILT) thickness. ILT thickness has been associated with AAA growth and rupture [13]. Like [12], we can incorporate this parameter using an INR with two output nodes, one for the AAA's lumen and one for its ILT. With this, we can obtain continuous representations of not only the evolving AAA's outer wall, but also the evolving lumen and ILT. This allows us to use ILT thickness as additional input feature for our GCN model. Other relevant parameters like age, smoking and cardiovascular diseases related to AAA growth can also be easily incorporated in the pipeline.

4.3.2 Additional Constraints

To ensure that an INR represents a proper SDF, recent studies impose additional constraints on the spatial gradients with respect to the input coordinates. The Eikonal regularization term has been widely used to learn SDFs, which constrains the norm of spatial gradients to be one at any location in the input domain. Ma et al. has introduced two additional constraints that enhance gradient consistency in the predicted SDFs. First, they introduced Neural-Pull, a method that constrains the directions of spatial gradients to pull surrounding 3D space onto the surface [37]. Specifically, they train a neural network to pull 3D locations to their closest points on the surface using the predicted SDF values and the gradient at the locations, both of which are computed by the network itself. More recently, Ma et al. also introduce a level set alignment loss to evaluate the parallelism of the SDF's level sets, which can be minimized to achieve better gradient consistency [38]. These loss terms can be applied as additional regularization terms in our loss function (Equation 2.3) to improve spatial gradient consistency in the predicted SDFs. Other recent work by Yang et al. shed light on the popular Eikonal loss [39]. In particular, they use partial differential equations (PDEs) to analyze the Eikonal loss, and show it can be unstable. They use geometric PDEs to propose a new loss regularization, i.e., second order derivative in the normal direction, that avoids over-regularization while stabilizing the Eikonal loss. This novel method called StEik allows for considering new network structures that are able to represent finer shape detail. In future work, we may consider these recent developments in our current implementation of the INRs.

4.3.3 Latent Space

To represent a single AAA shape, we optimize a single INR for a single patient. If we additionally condition the MLP on a low-dimensional, shape-specific latent vector, we are able to represent multiple AAA shapes using a single INR. This latent vector can be learned using an auto-encoder, which contains an encoder that compresses the input shape in a lower-dimensional latent vector. Alternatively, Park et al. proposed an encoder-less learning approach called DeepSDF, where a randomly initialized latent vector is mapped to a shape in the beginning of training, and the latent vectors are jointly optimized with the MLP weights during training [23]. During inference, the network weights are fixed, and an optimal latent vector is estimated.

Despite the lack of longitudinal image data, there is a wealth of image data available of individual AAA shapes. If we condition an MLP on a latent vector and jointly optimize them for a wealth of individual AAA shapes, we may obtain a smooth latent space of AAA shapes. A structured latent space may enable us to cluster AAA shapes based on geometrical parameters like AAA size and ILT thickness. Note that structuring the latent space requires additional regularization terms during optimization to ensure that related AAA shapes are close to each

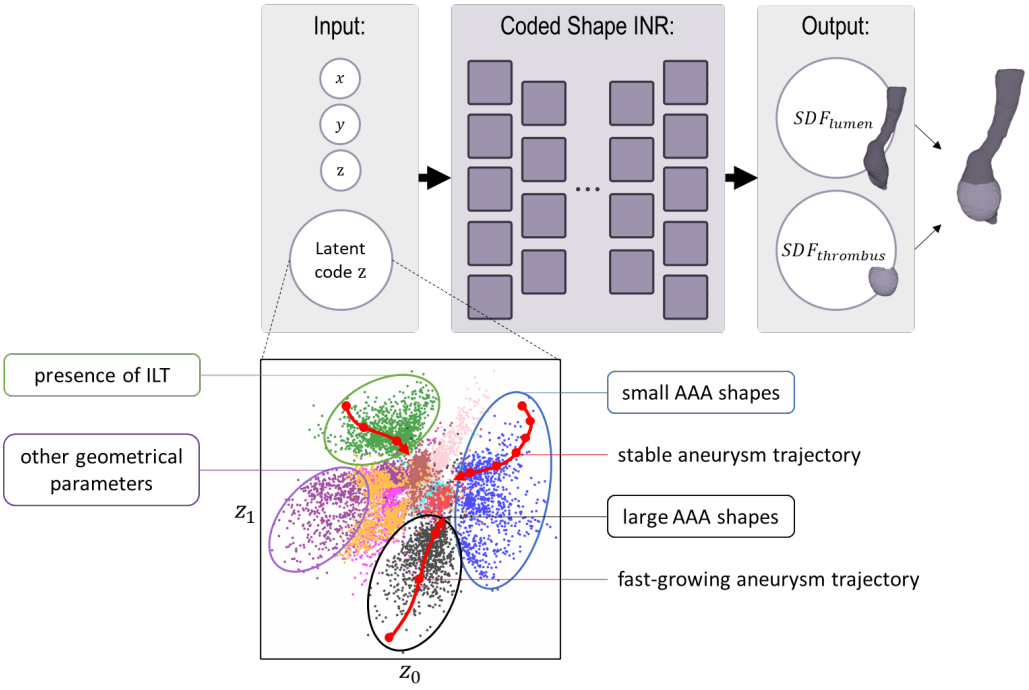


Figure 4.1: Schematic representation of proposed coded shape INR that is conditioned on a latent vector. During training, the latent vectors are jointly optimized with the MLP weights, resulting in a smooth AAA shape embedding space from which we can sample new synthetic shapes. Sampling a trajectory in this latent space generates a longitudinal synthetic dataset of an evolving AAA shape.

other in latent space. If we estimate the latent vectors of longitudinal AAA shapes of a single patient, we obtain a trajectory through latent space that corresponds to how the AAA shape evolves over time. These trajectories provide us with valuable information about how single AAA shapes move in latent space over time, which may allow us to distinguish between stable and fast-growing aneurysms. Another advantage of an optimized latent code is that we are able to generate realistic synthetic AAA shapes by sampling from this latent space. Since the latent space is smooth, we can sample an infinite amount of new AAA shapes. By sampling realistic trajectories in latent space, we may even generate longitudinal synthetic datasets of AAA shapes evolving over time. These synthetic datasets can be used as training data for our GCN to enhance robustness, generalization capabilities, and prediction performance. This proposed coded shape INR is similar to the work of Wiesner at al. where they jointly optimize an MLP and its latent space using a large set of 3D cell shapes evolving over time [25]. In Figure 4.1, we show a schematic representation of this proposed coded shape INR that is conditioned on a latent vector. Note that we also incorporated the two output nodes for the lumen and thrombus SDF, which together form the AAA's outer wall.

4.3.4 Uncertainty Quantification

A major barrier for adoption of a deep learning model in clinical practice can be mistrust of clinicians and patients in the model predictions. An important step in building their trust in deep learning is to provide a level of uncertainty associated with the model predictions. If we can quantify the uncertainty in the local AAA growth predictions made by the GCN model, clinicians and patients are equipped with valuable information regarding the utility of the model predictions; for example, in case of high predicted uncertainty, clinicians can ignore model predictions and rely more on their knowledge medical knowledge, experience and expertise.

In convolutional neural networks (CNNs), Monte-Carlo (MC) dropout has been a popular method for uncertainty quantification. Dropout is commonly used as a regularization technique during training to prevent overfitting. Dropout works by randomly dropping out nodes in the neural network during training. If used for uncertainty quantification, we only apply dropout during inference. In particular, multiple forward passes are performed with random dropout, and the variance in the predictions are a measure of uncertainty. Although CNNs and GCNs share similar network architecture, both with convolution, pooling and connection layers, the input data structures and convolution operations differ. Therefore, current dropout approaches for CNNs cannot directly be applied to GCNs. Instead of dropping out nodes in the neural network, random vertices are removed ('dropped out') from the input graph. If applied during inference, the variance of multiple predictions for a single input is a measure of uncertainty in the model's predictions. In future work, we may apply this approach to our GCN model to equip clinicians with a level of uncertainty associated with the model predictions. It is important to note that a low level of uncertainty is not always associated with a high accuracy of the model predictions.

4.3.5 Clinical Trial

Before our pipeline can be adopted in clinical practice, we should validate its performance in a clinical trial. In this trial, we could prospectively include AAA patients that will be monitored for the upcoming years. The objective is to do an up-scaled version of the experiments presented in Chapter 3, comparing the AAA shape predicted by our GCN model with the ground-truth AAA shape. For each scan instance, we can compare the maximum AAA diameter that is measured manually in the CTA image with the one extracted automatically from the AAA shape obtained using the segmentation model from our pipeline. Furthermore, using our optimized GCN network, we can predict local AAA growth on the surface over any future time period. If we predict the AAA shape from one scan instance to the next, we can quantitatively evaluate the GCN's performance in predicting local AAA growth by comparison to the ground truth AAA shapes. Metrics that can be used for this are (average) surface distance, difference in diameter along the AAA centerline, difference in maximum AAA diameter, Hausdorff distance and Dice similarity coefficient. In this clinical trial, we may also include other prediction models to compare the results for the same aneurysm dataset.

Prior to the clinical trial, we should carefully consider which metrics and corresponding threshold values are appropriate to determine when the prediction model is accurate enough to be adopted in clinical practice. In a study by Cayne et al., average measurement variability of maximum AAA diameter on the same CT scan is 4.0±5.1 mm [10]. If we are able to predict the maximum AAA diameter with a difference to ground-truth maximum AAA diameter below this inter-observer variability threshold, we can consider our prediction model as accurate enough to predict this parameter. Whereas this parameter is primarily used in clinical practice to assess rupture risk, we are also interested in determining whether the GCN model is accurate enough in predicting local AAA growth. In our GCN experiments, we observed that the average surface distances (ASDs) did not accurately reflect the performance of the GCN model in predicting local AAA growth. Instead, we should compare the diameter profiles; if the difference in the ground-truth and predicted diameter profile along the AAA centerline is less than the inter-observer variability, we can consider our prediction model as accurate enough to predict local AAA growth on the surface.

4.4 Conclusion

This study aimed to address the main research question: 'How can deep learning be used to predict personalized aneurysm growth based on longitudinal image data?'. To answer this question, we combine the continuity of INRs to obtain highly personalized, continuous models of 3D AAA shapes evolving over time with a GCN model that can learn regularizations from these representations to predict local AAA growth on the surface. With this, we are not only able to retrospectively assess but also predict AAA evolution over time. Ultimately, this pipeline has the potential to make clinical decisions regarding AAA surveillance more pro-active. Once the growth prediction model is accurate enough, it enables the clinicians to act on future events rather than react to the current situation, and may put the patient more at ease in case of a predicted stable aneurysm.

REFERENCES

[1] Jorik J Reimerink, Maarten J van der Laan, Mark J Koelemay, Ron Balm, and Dink A Legemate. Systematic review and meta-analysis of population-based mortality from ruptured abdominal aortic aneurysm. *Journal of British Surgery*, 100(11):1405–1413, 2013.

[2] PW Stather, D Sidloff, N Dattani, E Choke, MJ Bown, and RD Sayers. Systematic review and meta-analysis of the early and late outcomes of open and endovascular repair of abdominal aortic aneurysm. *Journal of British Surgery*, 100(7):863–872, 2013.

[3] George A Antoniou, Stavros A Antoniou, and Francesco Torella. Editor's choice–endovascular vs. open repair for abdominal aortic aneurysm: systematic review and meta-analysis of updated peri-operative and long term data of randomised controlled trials. *European Journal of Vascular and Endovascular Surgery*, 59(3):385–397, 2020.

[4] R Clement Darling. Ruptured arteriosclerotic abdominal aortic aneurysms: a pathologic and clinical study. *The American Journal of Surgery*, 119(4):397–401, 1970.

[5] H Glimåker, L Holmberg, A Elvin, O Nybacka, B Almgren, C-G Björck, and I Eriksson. Natural history of patients with abdominal aortic aneurysm. *European journal of vascular surgery*, 5(2):125–130, 1991.

[6] Anders Wanhainen, Fabio Verzini, Isabelle Van Herzeele, Eric Allaire, Matthew Bown, Tina Cohnert, Florian Dick, Joost van Herwaarden, Christos Karkos, Mark Koelemay, et al. Editor's choice–european society for vascular surgery (esvs) 2019 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *European Journal of Vascular and Endovascular Surgery*, 57(1):8–93, 2019.

[7] The UK Small Aneurysm Trial Participants. Mortality results for randomised controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. *The Lancet*, 352(9141):1649–1655, 1998.

[8] RC Darling, CR Messina, DC Brewster, and LW Ottinger. Autopsy study of unoperated abdominal aortic aneurysms. the case for early resection. *Circulation*, 56(3 Suppl):II161–4, 1977.

[9] Kevin P Conway, J Byrne, Matthew Townsend, and Isabel F Lane. Prognosis of patients turned down for conventional abdominal aortic aneurysm repair in the endovascular and sonographic era: Szilagyi revisited? *Journal of Vascular Surgery*, 33(4):752–757, 2001.

[10] Neal S Cayne, Frank J Veith, Evan C Lipsitz, Takao Ohki, Manish Mehta, Nick Gargiulo, William D Suggs, Alla Rozenblit, Zina Ricci, and Carlos H Timaran. Variability of maximal aortic aneurysm diameter measurements on ct scan: significance and methods to minimize. *Journal of vascular surgery*, 39(4):811–815, 2004.

[11] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[12] Dieuwertje Alblas, Christoph Brune, Kak Khee Yeung, and Jelmer M Wolterink. Going off-grid: Continuous implicit neural representations for 3d vascular modeling. In *Statistical Atlases and Computational Models of the Heart. Regular and CMRxMotion Challenge Papers: 13th International Workshop, STACOM 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Revised Selected Papers*, pages 79–90. Springer, 2023.

[13] Menno E Groeneveld, Jorn P Meekel, Sidney M Rubinstein, Lisanne R Merkestein, Geert Jan Tangelder, Willem Wisselink, Maarten Truijers, and Kak Khee Yeung. Systematic review of circulating, biomechanical, and genetic markers for the prediction of abdominal aortic aneurysm growth and rupture. *Journal of the American Heart Association*, 7(13):e007791, 2018.

[14] Karen López-Linares, Inmaculada García, Ainhoa García-Familiar, Iván Macía, and Miguel A González Ballester. 3d convolutional neural network for abdominal aortic aneurysm segmentation. *arXiv preprint arXiv:1903.00879*, 2019.

[15] Atefeh Abdolmanafi, Arianna Forneris, Randy D Moore, and Elena S Di Martino. Deep-learning method for fully automatic segmentation of the abdominal aortic aneurysm from computed tomography imaging. *Frontiers in Cardiovascular Medicine*, 9, 2022.

[16] Huan N Do, Ahsan Ijaz, Hamidreza Gharahi, Byron Zambrano, Jonguen Choi, Whal Lee, and Seungik Baek. Prediction of abdominal aortic aneurysm growth using dynamical gaussian process implicit surface. *IEEE Transactions on Biomedical Engineering*, 66(3):609–622, 2018.

[17] Liangliang Zhang, Byron A Zambrano, Jongeun Choi, Whal Lee, Seungik Baek, and Chae Young Lim. Intraluminal thrombus effect on the progression of abdominal aortic aneurysms by using a multistate continuous-time markov chain model. *Journal of International Medical Research*, 48(11):0300060520968449, 2020.

[18] Jordan B Stoecker, Kevin C Eddinger, Alison M Pouch, and Benjamin M Jackson. Image-derived geometric characteristics predict abdominal aortic aneurysm growth in a machine learning model. In *Functional Imaging and Modeling of the Heart: 11th International Conference, FIMH 2021, Stanford, CA, USA, June 21-25, 2021, Proceedings*, pages 35–45. Springer, 2021.

[19] Zhenxiang Jiang, Huan N Do, Jongeun Choi, Whal Lee, and Seungik Baek. A deep learning approach to predict abdominal aortic aneurysm expansion using longitudinal data. *Frontiers in Physics*, 7:235, 2020.

[20] Sekeun Kim, Zhenxiang Jiang, Byron A Zambrano, Yeonggul Jang, Seungik Baek, Sunkook Yoo, and Hyuk-Jae Chang. Deep learning on multiphysical features and hemodynamic modeling for abdominal aortic aneurysm growth prediction. *IEEE Transactions on Medical Imaging*, 42(1):196–208, 2022.

[21] Jonathan C Carr, Richard K Beatson, Jon B Cherrie, Tim J Mitchell, W Richard Fright, Bruce C McCallum, and Tim R Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 67–76, 2001.

[22] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, pages 641–676. Wiley Online Library, 2022.

[23] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[24] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*, 2020.

[25] David Wiesner, Julian Suk, Sven Dummer, David Svoboda, and Jelmer M Wolterink. Implicit neural representations for generative modeling of living cell shapes. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*, pages 58–67. Springer, 2022.

[26] David Lüdke, Tamaz Amiranashvili, Felix Ambellan, Ivan Ezhov, Bjoern H Menze, and Stefan Zachow. Landmark-free statistical shape modeling via neural flow deformations. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*, pages 453–463. Springer, 2022.

[27] Julian Suk, Pim de Haan, Phillip Lippe, Christoph Brune, and Jelmer M Wolterink. Mesh neural networks for se (3)-equivariant hemodynamics estimation on the artery wall. *arXiv preprint arXiv:2212.05023*, 2022.

[28] Dieuwertje Alblas, Marieke Hofman, Christoph Brune, Kak Khee Yeung, and Jelmer M Wolterink. Implicit neural representations for modeling of abdominal aortic aneurysm progression. *arXiv preprint arXiv:2303.01069*, 2023.

[29] Dawson-Haggerty et al. trimesh.

[30] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[31] H Gharahi, BA Zambrano, C Lim, J Choi, W Lee, and S Baek. On growth measurements of abdominal aortic aneurysms using maximally inscribed spheres. *Medical engineering & physics*, 37(7):683–691, 2015.

[32] Brian B Avants, Nicholas J Tustison, Michael Stauffer, Gang Song, Baohua Wu, and James C Gee. The insight toolkit image registration framework. *Frontiers in neuroinformatics*, 8:44, 2014.

[33] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.

[34] Pim De Haan, Maurice Weiler, Taco Cohen, and Max Welling. Gauge equivariant mesh cnns: Anisotropic convolutions on geometric graphs. *arXiv preprint arXiv:2003.05425*, 2020.

[35] Nicholas Sharp, Yousuf Soliman, and Keenan Crane. The vector heat method. *ACM Transactions on Graphics (TOG)*, 38(3):1–19, 2019.

[36] Anders Wanhainen, Fabio Verzini, Isabelle Van Herzeele, Eric Allaire, Matthew Bown, Tina Cohnert, Florian Dick, Joost van Herwaarden, Christos Karkos, Mark Koelemay, et al.

European society for vascular surgery (esvs) 2019 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *ACTA ANGIOLOGICA*, 28(3):69–146, 2022.

[37] Baorui Ma, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. *arXiv preprint arXiv:2011.13495*, 2020.

[38] Baorui Ma, Junsheng Zhou, Yu-Shen Liu, and Zhizhong Han. Towards better gradient consistency for neural signed distance functions via level set alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17724–17734, 2023.

[39] Huizong Yang, Yuxin Sun, Ganesh Sundaramoorthi, and Anthony Yezzi. Steik: Stabilizing the optimization of neural signed distance functions and finer shape representation. *arXiv preprint arXiv:2305.18414*, 2023.

A DIAMETER PROFILES

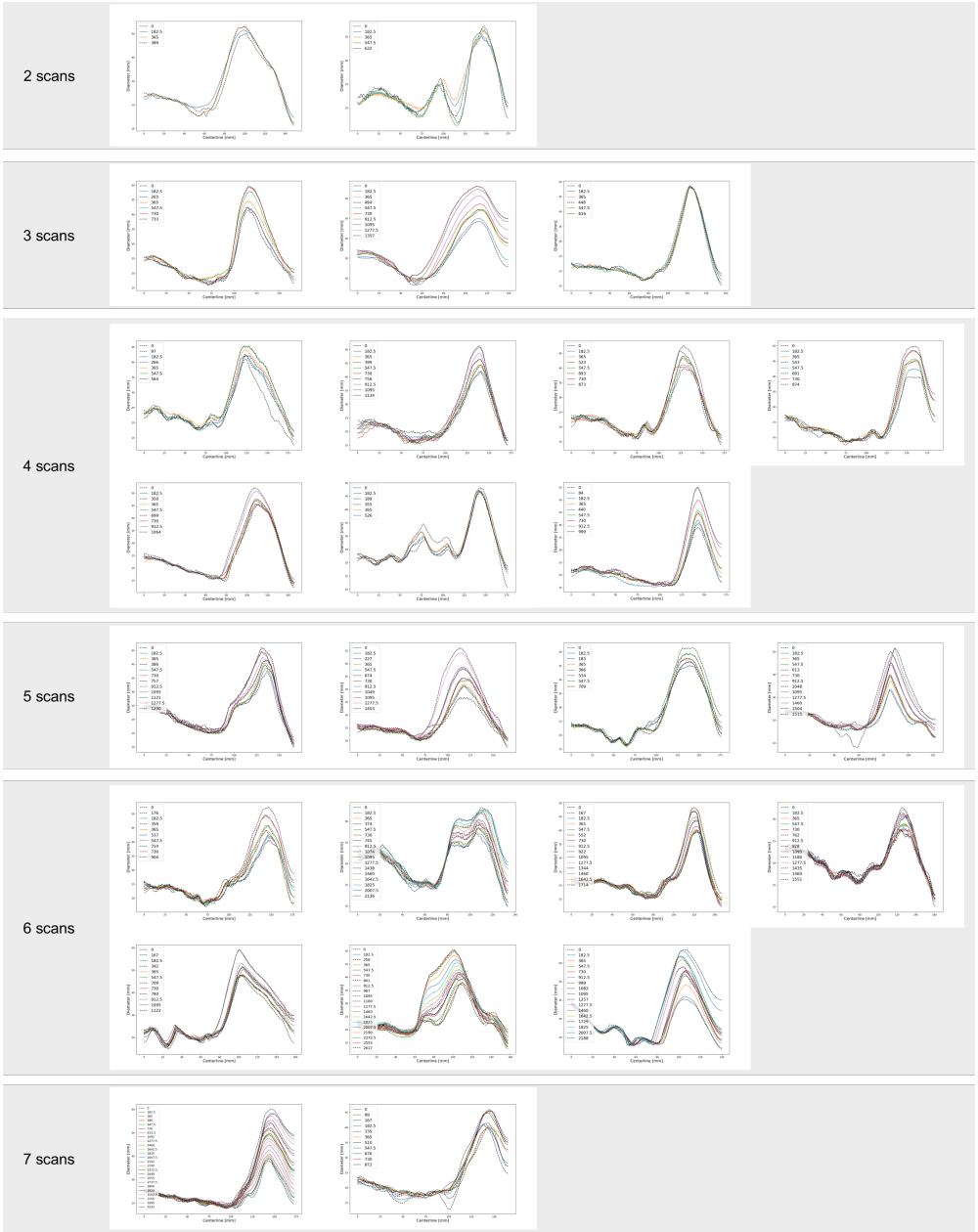
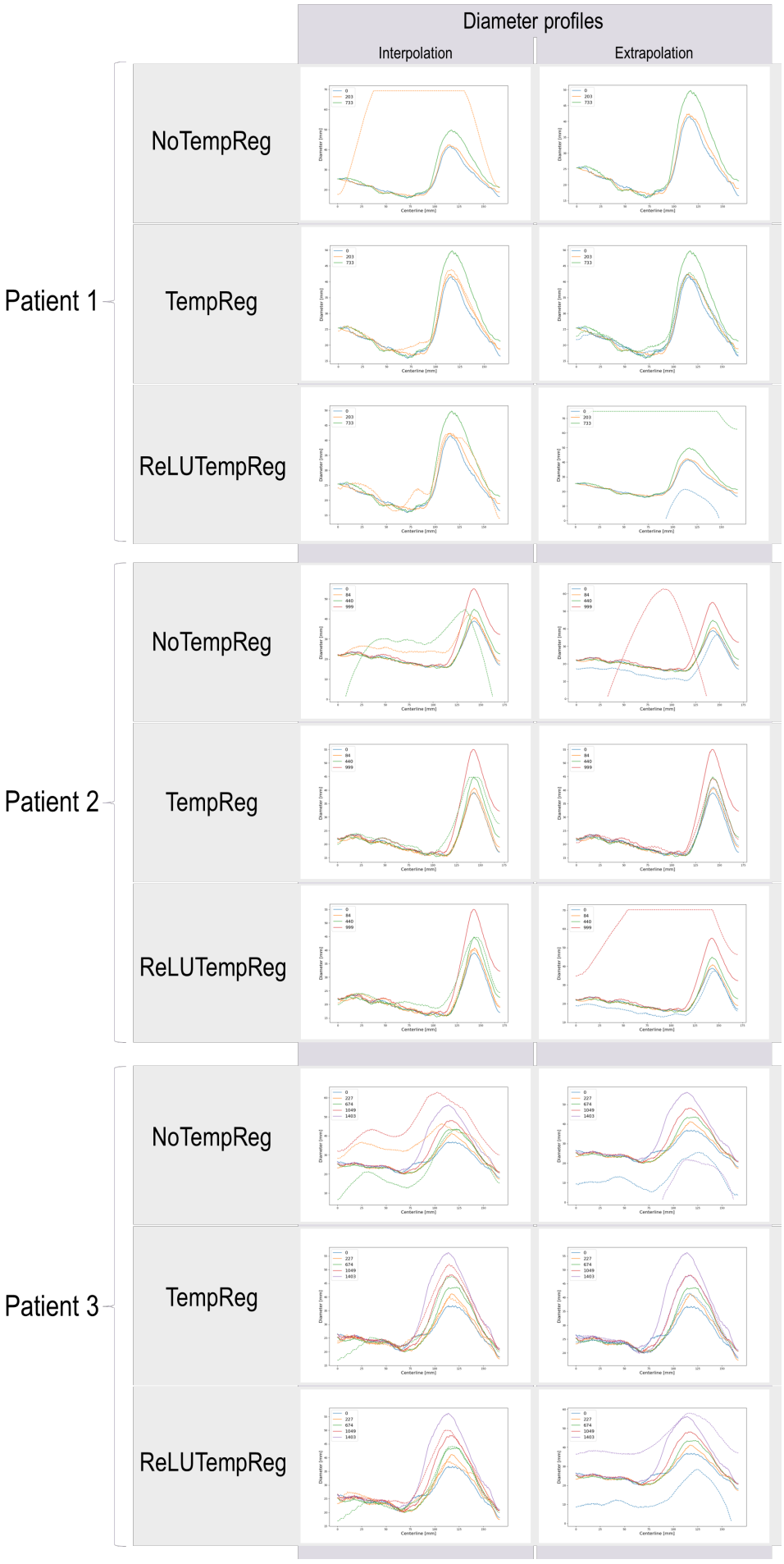


Figure A.1: Diameter profiles along axial slices for ground truth aorta segmentations (*black dashed lines*) and interpolated AAA surfaces extracted at time points spaced half a year apart (*colored solid lines*) for all 25 included patients. Each plot corresponds to a single patient, for which a single INR was optimized with TempReg based on point clouds from all available scan instances. The number of scan instances available during training is indicated in the left.


Diameter profiles
Interpolation
Extrapolation
Patient 1
NoTempReg
TempReg
ReLUTempReg
Patient 2
NoTempReg
TempReg
ReLUTempReg
Patient 3
NoTempReg
TempReg
ReLUTempReg

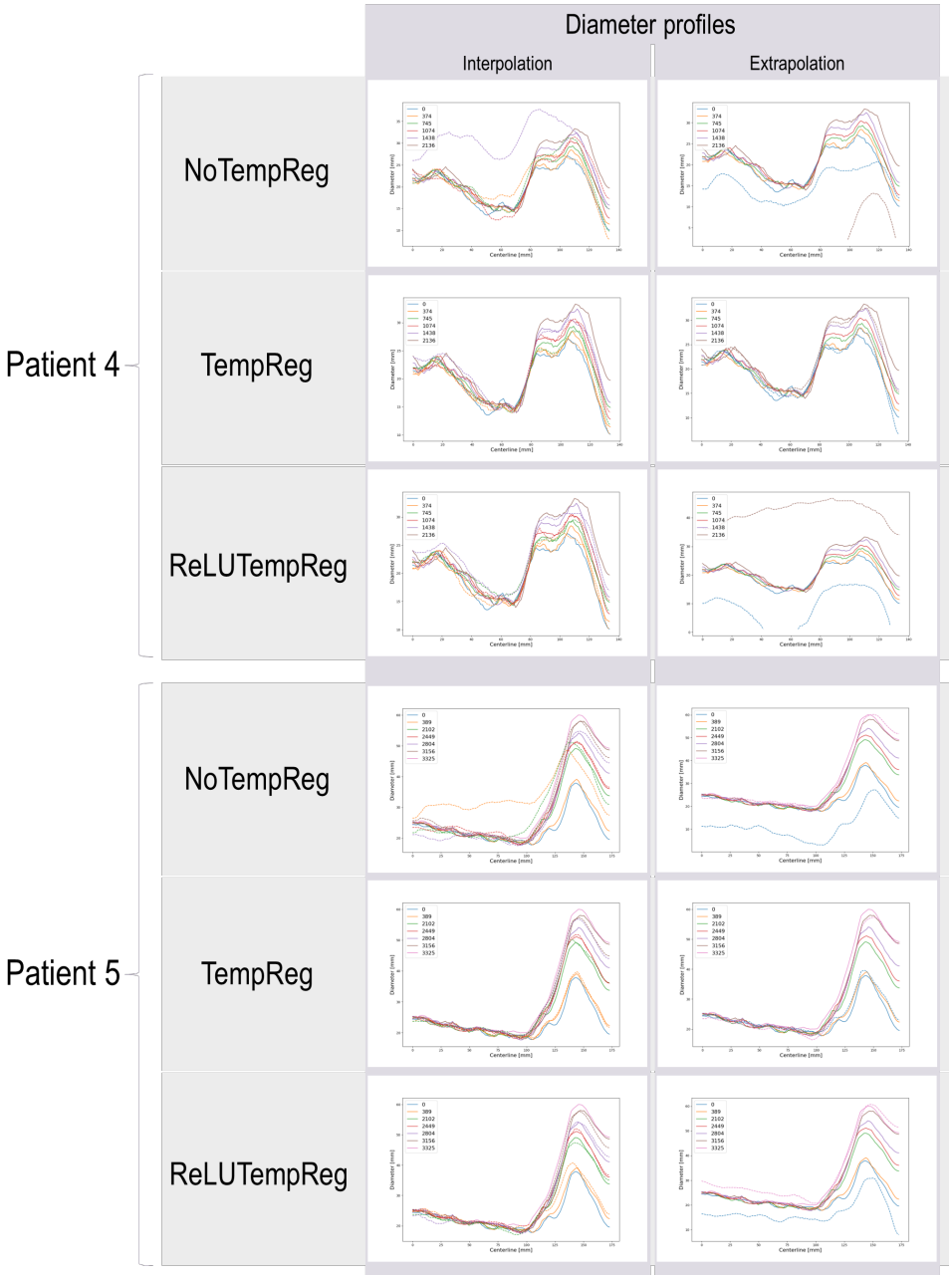


Figure A.2: Diameter profiles along axial slices for ground-truth aorta segmentations (*solid lines*) and inter- and extrapolated AAA shapes extracted at available scan instances (*dashed lines*) for five patients. For each patient, we compare the results for different temporal regularization strategies: NoTempReg (*top rows*), TempReg (*middle rows*), and ReLUTempReg (*bottom rows*).

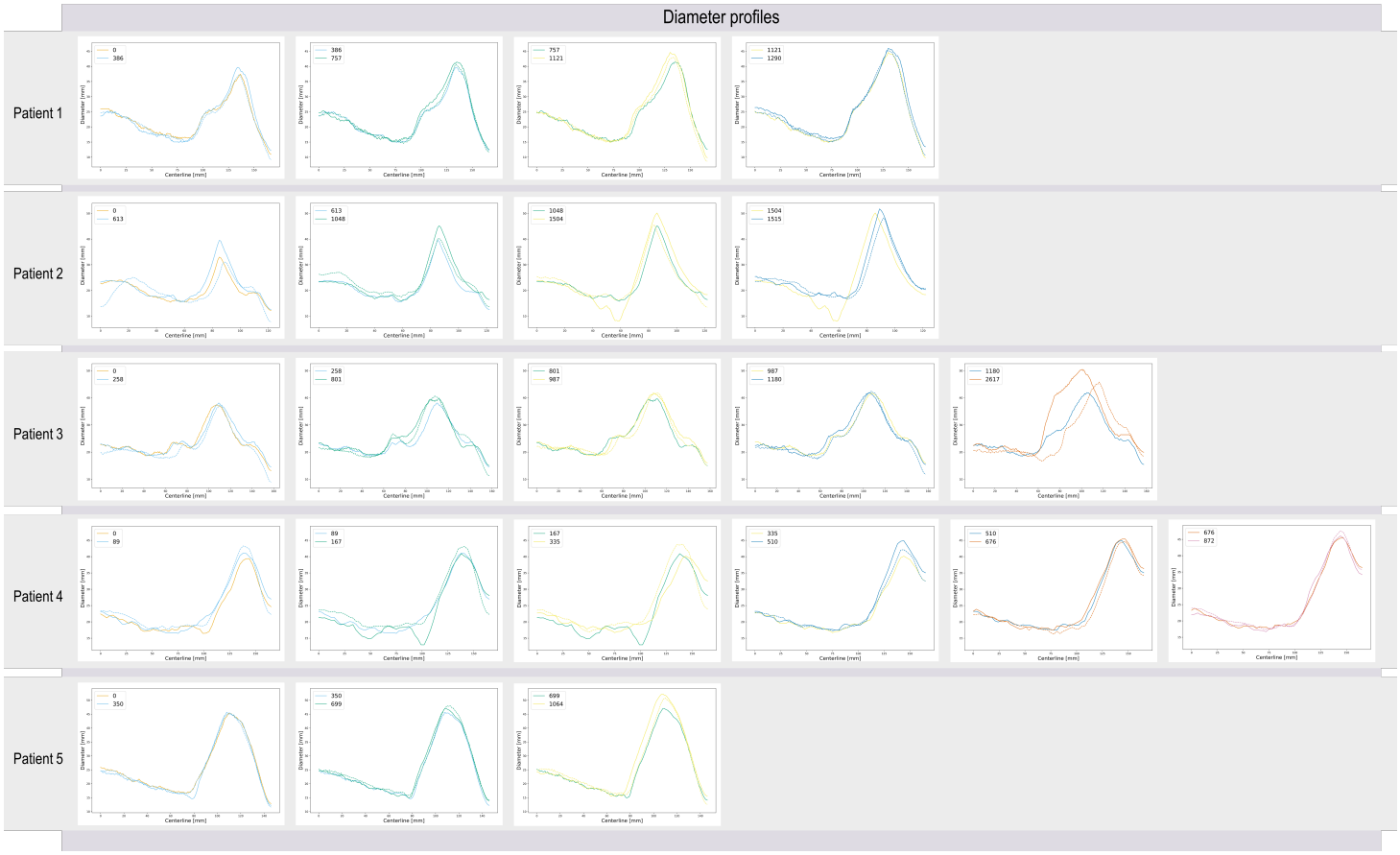


Figure A.3: Diameter profiles along the axial slices for all available ground aorta segmentations (*solid lines*) and predicted AAA shapes (*dashed lines*) for the five validation patients. In each plot, we compare three diameter profiles: the ground-truth and predicted diameter profiles at t_j, and the ground-truth diameter profile at t_{j-1} from which the prediction was made.