Inaugural – Dissertation

zur

Erlangung der Doktorwürde

der

Gesamtfakultät für Mathematik, Ingenieur- und Naturwissenschaften

der

Ruprecht-Karls-Universität Heidelberg

vorgelegt von

Holzwarth, Niklas Joshua, M.Sc.

aus Schwäbisch Hall

Tag der mündlichen Prüfung: _____

Context-aware clinical photoacoustic imaging

Supervisor: Prof. Dr. Lena Maier-Hein
Co-supervisor: Prof. Dr. Jürgen Hesser

Abstract

Photoacoustic imaging (PAI) is an emerging biomedical imaging modality that harnesses pulsed laser light to generate ultrasound waves through thermoelastic expansion, enabling high-resolution, non-invasive visualization of tissue structure and function at clinically relevant depths. By combining the optical contrast of molecular imaging with the spatial precision of ultrasound, PAI offers a unique capability to assess physiological and pathological processes in vivo. Its relatively low cost, safety, and imaging speed make it a highly promising tool for a wide range of clinical applications, including oncology and vascular medicine. The clinical adoption of PAI, however, has been limited by challenges in interpreting its data without sufficient spatial, temporal, and biophysical context. This thesis addresses this limitation by developing methods that incorporate contextual information across these dimensions, enabling more accurate and clinically meaningful PAI analysis.

The lack of spatial context was addressed with a framework for reconstructing three-dimensional (3D) volumes from sets of two-dimensional (2D) images. The central innovation of this approach lies in the use of an optical pattern that encodes spatial information through specific light-absorption characteristics. An extension of the pattern, adding fiducial markers, further enables the multimodal fusion of PAI with magnetic resonance imaging (MRI) and computed tomography (CT), thereby situating PAI within the established clinical imaging landscape. The lack of temporal context, addressed by pattern-based longitudinal registration of 3D PAI volumes, enables a more comprehensive assessment of disease status and progression. Third, a digital twin model was introduced to analyze unexpected clinical observations by disentangling physiological mechanisms from photoacoustic image formation processes. To demonstrate the broad applicability of these methods, they were validated in diverse clinical settings, with applications ranging from cancer therapy to vascular disease diagnosis. In a clinical study on peripheral artery disease, optical pattern-based PAI successfully detected ischemia and muscular heterogeneities, indicating benefits over conventional 2D approaches by combining spatial and temporal context. This thesis also presents the first evidence that PAI can non-invasively capture molecular changes induced by radiotherapy in patients with head and neck cancer. In this study, digital twin modeling further provided a mechanistic explanation for unexpected oxygenation

measurements, revealing that these anomalies arose from signal distortions in regions with low blood volume.

In conclusion, this work establishes the concept of context-aware PAI, integrating spatial and temporal, multimodal, and biophysical information to enhance both interpretability and clinical trust. By demonstrating feasibility in clinical studies, it outlines a pathway for translating context-aware PAI into routine medical practice.

Zusammenfassung

Photoakustische Bildgebung (PAI) ist ein neuartiges biomedizinisches Verfahren, das gepulstes Laserlicht zur Erzeugung von Ultraschallsignalen durch thermoelastische Expansion nutzt und eine hochaufgelöste, nicht-invasive Visualisierung von Gewebestruktur und -funktion in klinisch relevanten Tiefen ermöglicht. Durch die Kombination des optischen Kontrasts der molekularen Bildgebung mit der räumlichen Präzision des Ultraschalls bietet PAI eine einzigartige Möglichkeit, physiologische und pathologische Prozesse in vivo zu untersuchen. Ihre vergleichsweise geringen Kosten, Sicherheit und hohe Bildgebungsgeschwindigkeit machen sie zu einem vielversprechenden Werkzeug für zahlreiche klinische Anwendungen, insbesondere in der Onkologie und der Gefäßmedizin. Die klinische Etablierung von PAI wird jedoch bislang durch Herausforderungen bei der Interpretation der Bilddaten ohne ausreichenden räumlichen, zeitlichen und biophysikalischen Kontext eingeschränkt. Diese Dissertation adressiert diese Limitation, indem Methoden entwickelt wurden, diese kontextuellen Informationen zu integrieren und dadurch eine präzisere und klinisch aussagekräftigere PAI-Analyse ermöglichen.

Der fehlende räumliche Kontext wurde durch die Entwicklung eines Systems zur Rekonstruktion dreidimensionaler (3D) Volumina aus zweidimensionalen (2D) Schichtbildern adressiert. Die zentrale Innovation liegt in der Verwendung eines optischen Musters, das räumliche Informationen über spezifische Lichtabsorptionseigenschaften kodiert. Eine Erweiterung dieses Musters um Referenzmarker ermöglicht zudem die multimodale Fusion von PAI mit Magnetresonanztomographie (MRI) und Computertomographie (CT) und positioniert PAI damit innerhalb der etablierten klinischen Bildgebungslandschaft. Der fehlende zeitliche Kontext wurde durch eine optische Muster-basierte longitudinale Registrierung von 3D-PAI-Volumina adressiert, die eine umfassendere Beurteilung des Krankheitsverlaufs und -fortschritts ermöglicht. Drittens wurde ein *digital twin*-Modell eingeführt, um unerwartete klinische Beobachtungen zu analysieren und physiologische Mechanismen von Limitierungen der photoakustischen Bildentstehung zu trennen.

Zur Demonstration der breiten Anwendbarkeit dieser Methoden wurden sie in verschiedenen klinischen Szenarien validiert, von der Krebstherapie bis zur Diagnostik vaskulärer Erkrankungen. In einer klinischen Studie zur peripheren arteriellen Verschlusskrankheit konnte die optische Muster-basierte PAI erfolgreich Ischämien und

muskuläre Heterogenitäten detektieren und zeigte dabei Vorteile gegenüber herkömmlichen 2D-Ansätzen durch die Kombination räumlicher und zeitlicher Kontextinformation. Darüber hinaus liefert diese Arbeit erstmals den Nachweis, dass PAI nicht-invasiv molekulare Veränderungen erfassen kann, die durch Strahlentherapie bei Patienten mit Kopf-Hals-Tumoren hervorgerufen werden. In dieser Studie ermöglichte das *digital twin*-Modell eine mechanistische Erklärung für unerwartete Oxygenierungsmessungen und zeigte, dass diese Anomalien auf Signalverzerrungen in Regionen mit geringem Blutvolumen zurückzuführen sind.

Zusammenfassend etabliert diese Arbeit das Konzept der kontextbasierten PAI, die räumliche und zeitliche, multimodale und biophysikalische Informationen integriert, um sowohl die Interpretierbarkeit als auch das klinische Vertrauen zu stärken. Dabei zeigt der Nachweis der Machbarkeit in klinischen Studien einen Weg für die Umsetzung der kontextbasierten PAI in die medizinische Praxis auf.

Acknowledgments

Over 7.5 years have passed since I first started exploring photoacoustic imaging during my bachelor's in physics. Sometimes, when I look back, it feels like an eternity; other times, I wonder how fast it all went by. It probably has something to do with the theory of relativity, but I'm not an expert.

Now, the time has come for the next big milestone, and I'm very grateful to everyone who supported me along the way: in work (including proofreading this thesis), life, and the ever-tricky balance between the two. I can't name you all, even if I tried - thank you!

Since I already failed successfully at keeping the thesis short, I think I can spare some extra words here. Without doubt, I would not have started a PhD if it hadn't been for my supervisor, Lena Maier-Hein. I originally only planned to continue my master's project as a research scientist, but she encouraged me early on and supported my academic path from the very beginning until today. She believed in me and in my work, providing invaluable guidance, help, and constructive feedback. The positive and highly motivating work environment she fosters, filled with young, enthusiastic scientists, was and still is a pleasure to be part of. Thank you.

Many thanks also go to the entire CAMI team, or IMSY, as it was rebranded not so recently, and the MIC group for filling that environment with life, from coffee breaks and retreats to Christmas parties and much more. And, of course, to *THE office* for their support, including helping me navigate the sometimes stormy waters of German bureaucracy.

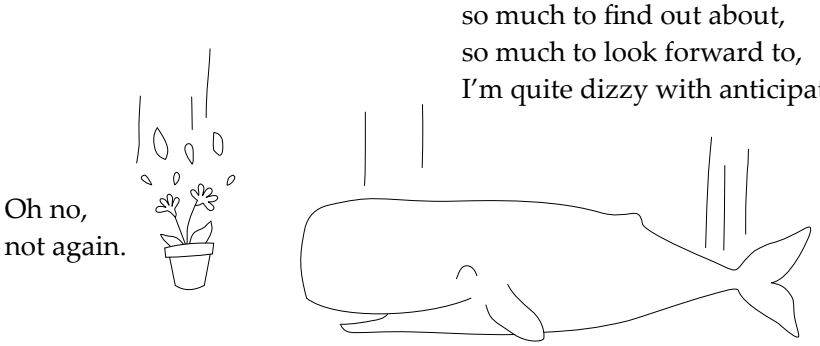
Special thanks to all current and former Biophotonics members, and especially the ISPAI team. Thomas, who introduced me to the department and, together with Janek, supervised my work in the early years. And thank you, Alex, for taking over, for always having an open ear, your thoughtful feedback, and the outstanding pizza.

Thanks to Kris, always bringing crystal-clear insights and even sharper puns, Jan-Hinrich, Melanie, Tom, CJ, and Marcel, the best colleagues, office partners, conference buddies, Bergfestler, and much more, that I could have wished for. And to all other former ISPAI students, especially Patricia and Sebastian, whom I had the pleasure to supervise during their master's theses.

My thanks also go to my thesis advisory committee. Jürgen Hesser and Jörg Peter provided not only critical and constructive feedback at the TAC meetings but were always open for helpful discussions and input. I would also like to thank the clinical collaborators at the University Clinics of Heidelberg and Erlangen, and the department of veterinary sciences at the LMU Munich. Thank you for your support!

Last but most definitely not least, thanks to the Le BinPhys for the time in Heidelberg together. To my family: thank you for your unwavering support, not least for babysitting during the writing phase. And Anne, you always had my back over the last five years. You made the writing phase of this thesis as relaxed as it could possibly be. If we weren't already married, I would propose to you every single day. And of course, Frieda, you spark joy every time I get to spend time with you. I'm truly looking forward to all the adventures ahead with you three ♡.

What is there left to say?



The Hitchhiker's Guide to the Galaxy
Douglas Adams (1979)

Contents

Lists

List of abbreviations

2D	two-dimensional
3D	three-dimensional
ABI	ankle-brachial index
API	application programming interface
AUROC	area under the receiver operating characteristic curve
BFGS	Broyden–Fletcher–Goldfarb–Shanno
BVF	blood volume fraction
BVM	German Conference on Medical Image Computing
CE	Conformité Européenne
CF	cystic fibrosis
CI	confidence interval
cINN	conditional invertible neural network
CNN	convolutional neural network
CPU	central processing unit
CT	computed tomography
DAS	delay-and-sum
DKFZ	German Cancer Research Center
DNA	deoxyribonucleic acid
DoF	degrees of freedom
DSC	Dice similarity coefficient
FAU	Friedrich-Alexander-Universität
FDA	Food and Drug Administration
FFT	fast Fourier transform
FLE	fiducial localization error
FN	false negative
FOV	field of view
FP	false positive
FRE	fiducial registration error
GUI	graphical user interface
Hb	deoxygenated hemoglobin
HbO_2	oxygenated hemoglobin
HN	head and neck
HU	Hounsfield units
HV	healthy volunteer
IMU	inertial measurement unit
IPASC	International Photoacoustic Standardisation Consortium
IPCAI	Information Processing in Computer-Assisted Interventions
IQR	interquartile range
ISO	International Organization for Standardization
ISPAI	Intelligent Systems in Photoacoustic Imaging
LED	light-emitting diode
LMM	linear mixed model
LOO	leave-one-out
LR^+	positive likelihood ratio
MICCAI	Medical Image Computing and Computer-Assisted Intervention
MITK	Medical Imaging Interaction Toolkit
MRI	magnetic resonance imaging
MSOT	multispectral optoacoustic tomography
NCT	National Center for Tumor Diseases
PA	photoacoustic
PAD	peripheral artery disease
PAI	photoacoustic imaging
PCA	principal component analysis
PET*	polyethylene terephthalate
PET	positron emission tomography
PPV	positive predictive value
PRF	pattern reference frame
RAM	random access memory
RBE	relative biological effectiveness
RECIST	response evaluation criteria in solid tumors
RMS	root-mean-square
RMSE	root mean squared error
ROI	region of interest

RT	radiotherapy
RTE	radiative transfer equation
SIMPA	Simulation and Image Processing for Photonics and Acoustics
SNR	signal-to-noise ratio
sO_2	oxygen saturation
SVD	singular value decomposition
tHb	total hemoglobin
TN	true negative
TMR	tumor-to-muscle ratio
TP	true positive
TRE	target registration error
TRL	technology readiness level
t-SNE	t-distributed stochastic neighbor embedding
UML	unified modeling language
UMAP	uniform manifold approximation and projection for dimension reduction
US	ultrasound
UV	ultraviolet

List of figures

List of tables

List of algorithms

- Part I: Fundamentals -

1 Introduction

1.1 Motivation

Ultrasound (US) has been instrumental in the evolution of modern healthcare, forming a fundamental component of the diagnostic repertoire alongside and together with X-ray, computed tomography (CT), and magnetic resonance imaging (MRI). Originating from echolocation technologies developed during World War II, ultrasound has become an affordable, non-invasive, and portable imaging modality, enabling real-time visualization of anatomy in routine clinical care [1].

US imaging operates on the principle of *sound in – sound out*. Piezoelectric elements in the US transducer emit ultrasound waves in the megahertz range and measure the reflected signals caused by changes in acoustic impedance between different tissue types [2]. As early as the late 1950s, Japanese researchers discovered that the frequency of these reflected waves is proportional to blood flow velocity, paving the way for Doppler ultrasound, which is now a standard diagnostic tool for assessing blood flow in conditions such as peripheral artery disease (PAD) [3]. These early developments already reflect a growing clinical demand for imaging modalities that go beyond anatomical visualization and provide insight into functional (metabolic) processes [4].

Photoacoustic imaging (PAI), also known as optoacoustic imaging, addresses this need. As "arguably the most exciting biomedical imaging technique of the decade" [5], PAI is based on the photoacoustic (PA) effect, first discovered by Alexander Graham Bell in 1880. It involves the generation of US waves via thermoelastic expansion caused by the absorption of short-pulsed laser light [5]: *light in – sound out*. Like optical imaging, PAI can employ multispectral measurements, using various laser wavelengths, to estimate wavelength-dependent light absorption and thereby infer molecular tissue composition. Relevant optical biomarkers include oxygenated and deoxygenated hemoglobin (carried by red blood cells), water, and lipids, as well as derived parameters such as oxygen saturation (sO_2). These biomarkers are key to detecting, monitoring, and characterizing various diseases. As a result, PAI holds transformative potential across many high-impact clinical domains, including oncology and cardiovascular disease imaging with the potential to impact "the entire palette of modern biology" [5].

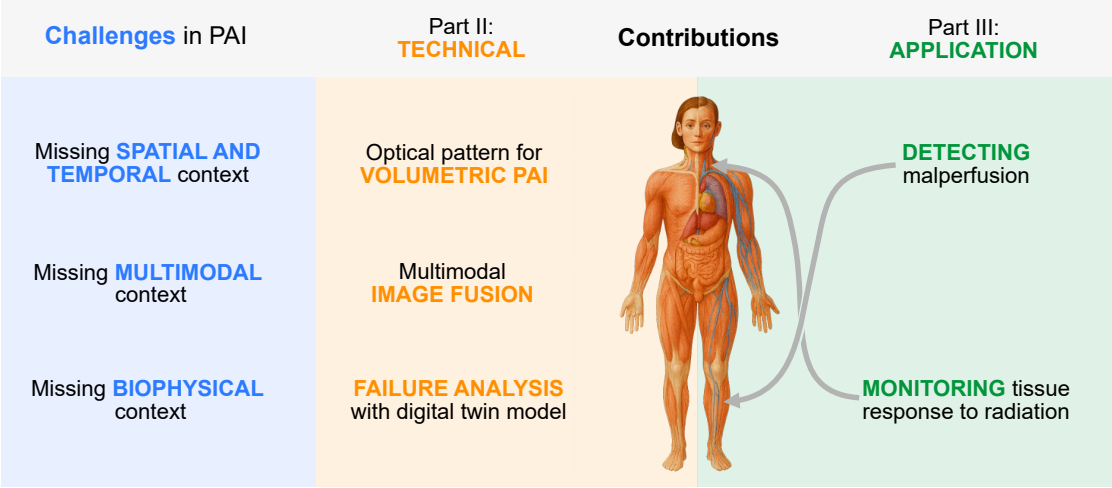


Figure 1.1: The figure gives an overview of selected challenges *missing context* in clinical photoacoustic imaging (PAI) (blue). The matching technical contributions (Part II) of this thesis are highlighted in orange, whereas the clinical applications (Part III) are displayed in green.

However, despite its promise, PAI remains underexplored, and several barriers hinder its clinical translation [6]. These include: (1) unsolved challenges in quantitative imaging, (2) lack of standardization, and (3) the need for clear identification of high-value clinical applications. With this thesis, I aim to help bridge the translational gap by proposing *context-aware clinical photoacoustic imaging* (Fig. 1.1), a concept that integrates three complementary dimensions of context to make this emerging modality clinically actionable:

1) **Spatial and temporal context**: It has been shown that reducing the assessment of complex anatomical changes to measurements in two-dimensional (2D) image space risks losing important information needed to draw robust clinical conclusions. A prominent example is the use of the response evaluation criteria in solid tumors (RECIST), which can lead to substantial variability and uncertainty in assessing tumor burden [7, 8]. For PAI, where functional signals may vary across space and time, this lack of spatial context of conventional 2D PAI is particularly limiting. Embedding spatial and temporal context through volumetric and longitudinal three-dimensional (3D) PAI provides a more comprehensive and reproducible basis for clinical interpretation.

2) **Multimodal context**: Different imaging modalities provide complementary information on anatomy, function, and molecular processes. Recent work emphasizes

that PAI achieves great value when combined with established modalities, such as MRI and CT, enabling direct anatomical grounding while contributing unique optical contrast for vascular and functional insights [9]. Such multimodal fusion strengthens clinical trust and situates PAI within existing workflows.

3) **Biophysical context**: Quantitative interpretation of PAI signals is fundamentally challenged by physical effects in image formation, including light fluence variations and spectral coloring, which can distort chromophore quantification and impede reliable oxygenation readouts [10]. Understanding the interplay between physiology and PA image formation through digital-twin modeling provides the necessary biophysical context to separate biological changes from methodological limits.

1.2 Objectives

The overarching objective of this thesis is to advance the clinical translation of PAI by establishing a framework for context-aware photoacoustic imaging. Specifically, the thesis is directed toward overcoming the following limitations:

1) the restriction of conventional PAI to a 2D tomographic slice, limiting spatial context and reproducibility;

2) the need for comparative imaging to improve anatomical interpretability, motivating the development of multimodal registration techniques;

3) the limited understanding of the interaction between biological mechanisms and PAI signal generation, which hinders accurate interpretation of functional imaging results.

1.2.1 Technical objectives

To address these limitations and to demonstrate the feasibility of robust, context-aware PAI in diverse clinical applications, three main technical objectives are defined:

Technical objective 1: Framework for volumetric PAI in clinical settings
The first objective is the development of a unified framework that enables reproducible and robust PA image acquisition by providing 3D spatial and temporal context for clinical application. A central focus lies on methods suitable for in vivo imaging, as

well as on validating the individual components of the framework to ensure clinical feasibility and robustness.

Technical objective 2: Marker-based multimodal 3D image registration
The second objective is to establish multimodal context by enabling the integration of 3D PAI with established imaging modalities such as MRI and CT. The goal is to provide direct anatomical reference for functional PAI data, thereby complementing existing clinical imaging, which currently lacks this type of functional information.

Technical objective 3: Digital twin–based concept for failure analysis in clinical PAI
The third objective is to strengthen the biophysical context of clinical PAI by elucidating how physiological and physical mechanisms manifest in PA signal generation. This objective aims to provide insights into the mechanisms and constraints underlying signal formation observed in clinical practice, and to establish a systematic basis for identifying and explaining unexpected measurement outcomes.

1.2.2 Clinical objectives

This thesis focuses on two clinical applications chosen to demonstrate the broad applicability of context-aware PAI in distinct disease settings: tissue response to radiotherapy in head and neck cancer, and diagnosis of PAD. In both cases, PAI has the potential to fill critical gaps in diagnosis and monitoring by enhancing conventional ultrasound with quantitative tissue composition and biomarker information, most notably sO_2, which is crucial for identifying hypoxia in tumors and ischemia in cardiovascular disease.

Clinical objective 1: Photoacoustic imaging for monitoring radiotherapy treatment response in head and neck tumors
In radiotherapy, there is currently no means of providing early treatment-response feedback in clinical routine, although such information would be essential for adaptive dose management. The first clinical objective of this work is therefore to evaluate whether PAI can provide longitudinal feedback on biological tissue responses to radiotherapy in patients with head and neck cancer. By acquiring PAI data at multiple time points throughout the course of treatment, this study aims to characterize signal variations

that reflect radiotherapy-induced physiological changes, thereby offering the potential for earlier, non-invasive assessment of treatment efficacy.

Clinical objective 2: Volumetric photoacoustics for detecting peripheral artery disease
A large proportion of PAD cases remain asymptomatic and underdiagnosed, as current low-cost diagnostic procedures are particularly unreliable in early stages. Yet, early detection is crucial, since it allows for timely, non-invasive interventions. The second clinical objective is therefore to demonstrate the feasibility of 3D PAI for in vivo detection of PAD (claudication) and the assessment of muscle tissue heterogeneities. This includes the analysis of longitudinal 3D measurements acquired before and after a standardized exercise protocol, aiming to identify functional impairments that remain invisible to conventional diagnostic methods.

1.3 Outline

The thesis is organized into four main parts. Part I provides the fundamental background necessary to understand the core principles of clinical 3D PAI, along with a summary of the current state of the art in the field. Part II focuses on the technical contributions of this work. Sections 5, 6, and 7 in this part align with the defined objectives, covering the hardware, algorithms, and implementation of an optical pattern approach for 3D PAI, pattern-based multimodal PAI, and digital twin-based failure analysis. Part III addresses the clinical applications of PAI, demonstrating how the newly developed techniques can be deployed to tackle specific diagnostic and monitoring challenges namely, leveraging biophysical and temporal context by monitoring tissue response to radiotherapy, as well as using spatial and temporal context provided by the optical pattern to detect PAD. Each section in Parts II and III is subdivided into materials and methods, validation and/or results, and discussion subsections. Part IV concludes the thesis by summarizing the key contributions, providing a closing discussion, and offering conclusions and an outlook on future research directions.

2 Background

In this chapter, the fundamental background knowledge required for understanding and implementing this work is presented. It begins with an introduction to the principles of PAI (Sec. 2.1), providing a deep understanding of how light interacts with biological tissue and how functional tissue information can be derived from these interactions.

Section 2.2 then covers image registration, with a focus on rigid transformations. This includes aligning image pairs in terms of position, orientation, and time, which is essential for multimodal and longitudinal imaging studies.

The following section (Sec. 2.3) offers a brief overview of key concepts in statistics and data science. These tools are critical for analyzing and interpreting results in clinical research, especially in the context of hypothesis testing and data-driven modeling.

Finally, Section 2.4 discusses the clinical background relevant to this work. It focuses on selected application areas where photoacoustic imaging shows significant promise: head and neck radiotherapy and peripheral artery disease.

2.1 Photoacoustic imaging

PAI is a biomedical imaging modality that offers a low-cost, non-invasive, and radiation-free method for visualizing functional tissue parameters, such as sO_2, with spatial resolution in depth. It is fundamentally characterized by a *light in – sound out* mechanism, where pulsed optical energy is absorbed by tissue chromophores, leading to localized thermoelastic expansion and the generation of broadband ultrasound waves.

This section draws heavily on foundational concepts presented in the textbook *Biomedical Optics: Principles and Imaging* by Wang and Wu [11]. It begins with an overview of light-tissue interactions in the near-infrared window, which is particularly relevant due to the relatively low absorption and scattering properties of biological tissues in this spectral range. The discussion then progresses to the principles of image generation, including the detection of induced acoustic signals and the reconstruction of spatially resolved absorption maps.

Subsequently, the focus shifts to the estimation of functional parameters using spectral unmixing techniques, which exploit the wavelength-dependent optical absorption characteristics of endogenous chromophores. Finally, the section concludes by introducing the basic principles behind synthetic image formation.

2.1.1 Light-tissue interaction

Light, defined as the visible portion of the electromagnetic spectrum, interacts with biological tissue in three primary ways: reflection/refraction, scattering, and absorption. Reflection and refraction occur at the interface between two media with differing refractive indices, governed by Snell's law [12]. If the angle of incidence exceeds the critical angle, total internal reflection occurs.

Once light enters the tissue, scattering dominates in the wavelength range of interest. Biological tissue is thus considered a scattering, or turbid, medium [11]. Scattering refers to the change in direction of photons due to interaction with particles. This phenomenon is described by Mie theory or, when the scatterers are significantly smaller than the wavelength, by Rayleigh theory [11].

In tissue, the mean free path length of a photon, the average distance it travels before a scattering event, is approximately 0.1 nm. The likelihood of scattering per unit length is described by the scattering coefficient μ_s. The directionality of scattering is characterized by the anisotropy factor g, which reflects the degree of forward-directed scattering.

After multiple scattering events, and in the absence of absorption, light may exit the tissue either on the same side (diffuse reflectance) or on the opposite side (transmission). These phenomena form the basis of various biomedical imaging modalities. For instance, hyperspectral imaging relies on diffuse reflectance, while transmission pulse oximetry is based on transmitted light. In both cases, the spectral dependence of the measured light is strongly influenced by the tissue composition, particularly by its wavelength-dependent absorption.

Absorption is a process in which a photon's energy excites an electron from its ground state to a higher energy state. Upon relaxation from this excited state, the electron may emit another photon (fluorescence) or dissipate energy as vibrational heat (nonradiative relaxation). The quantum efficiency describes the ratio between the number of emitted and absorbed photons.

Analogous to scattering, the likelihood of absorption per unit length is described by the absorption coefficient μ_a, which is the reciprocal of the mean free absorption path length and typically in the range of 10–100 nm in tissue [11].

Radiative transfer In general, light–tissue interaction in biological media can be analytically described by the radiative transfer equation (RTE), also known as the Boltzmann transport equation, under the assumptions that polarization, coherence,

and nonlinear effects are negligible. The model assumes a time-invariant scenario with spatially varying optical coefficients μ_a, μ_s, and g, and considers only elastic scattering.

The RTE can be derived from the fundamental physical principle of energy conservation within a closed system, applied to an infinitesimal differential volume element dV and a solid angle element $d\Omega$. It accounts for the balance of energy carried by photons traveling in a given direction within the medium.

Four distinct terms contribute to the differential change in radiative energy dP. These terms represent either a loss $(-)$ or a gain $(+)$ of energy in the system, and collectively form the basis of the RTE formulation:

- $-$ divergence: dP_{div}
- $-$ extinction due to absorption and scattering: dP_{ext}
- $+$ scattering into $d\Omega$: dP_{sca}
- $+$ energy source within dV: dP_{src}

This leads us to (Eq. 5.22 in [11]):

$$dP = -dP_{\text{div}} - dP_{\text{ext}} + dP_{\text{sca}} + dP_{\text{src}} \tag{2.1}$$

and by substituting (Eq. 5.13, 5.14, 5.19-5.21, and 5.22 from [11]):

$$\begin{aligned}\frac{\partial L(\vec{r},\hat{s},t)/c}{\partial t} &= -\hat{s}\cdot\nabla L(\vec{r},\hat{s},t) - \mu_t L(\vec{r},\hat{s},t) \\ &\quad + \mu_s \int_{4\pi} L(\vec{r},\hat{s}',t) P(\hat{s}'\cdot\hat{s})\, d\Omega' + S(\vec{r},\hat{s},t),\end{aligned} \tag{2.2}$$

we end up with the RTE where

- L/c is the propagating energy per volume and solid angle,
- $\hat{s}$ the propagation direction,
- $\mu_t = \mu_a + \mu_s$ the probability of extinction, and
- S the energy provided by source.

Due to the high dimensionality and complexity of light–tissue interactions, solving the RTE numerically is computationally intensive and often impractical for large-scale simulations in heterogeneous media. To facilitate practical modeling, a common

simplification assumes isotropic radiance after sufficient scattering and that absorption is much smaller than scattering, i.e., $\mu_a \ll \mu_s$. Under these conditions, the RTE reduces to what is known as the diffusion approximation. This leads to the diffusion equation, which provides a more tractable and analytically solvable description of light transport in highly scattering media. The resulting expression corresponds to Eq. 5.77 in [11]:

$$\frac{\partial \phi(\vec{r},t)}{c\,\partial t} + \mu_a \phi(\vec{r},t) - \nabla \cdot [D \nabla \phi(\vec{r},t)] = S(\vec{r},t), \tag{2.3}$$

with the fluence rate $\Phi(\vec{r},t)$, the isotropic source $S(\vec{r},t)$ and the diffusion coefficient D:

$$D = \frac{1}{3(\mu_a + \mu_s \cdot (1-g))}. \tag{2.4}$$

2.1.2 Near-infrared window

Professor Frans Jöbsis, a pioneer in biomedical optics [13], is the first to describe the near-infrared window [14]. This window originally defines the wavelength range from 700 nm to 1300 nm, with slight variations depending on the referenced source.

The near-infrared window is characterized by the low absorption of water, which allows for relatively long photon penetration depth through biological tissue due to the extended mean free path length. This optical property enables a range of biomedical optics applications [15] which rely on infrared transmission, reflection or absorption characteristics, including PAI.

Strictly speaking, the term *light* refers only to wavelengths within the human visible electromagnetic spectrum, which has been shown to extend under optimal conditions from approximately 310 nm into the near-infrared range up to 1100 nm [16]. However, for simplicity within this thesis, the term *light* extends to include the full wavelength range of the near-infrared window, with a particular interest in red and near-infrared light spanning from 660 nm to 1300 nm.

2.1.3 Photoacoustic image formation

"In bringing before you some discoveries made by Mr. Sumner Tainter and myself [Alexander Graham Bell], which have resulted in the construction of apparatus for the production and reproduction of sound by means of light [...]" - Alexander Graham Bell 1880, Bosten (USA) [17].

Although the PA effect was first discovered by Alexander Graham Bell in 1880,

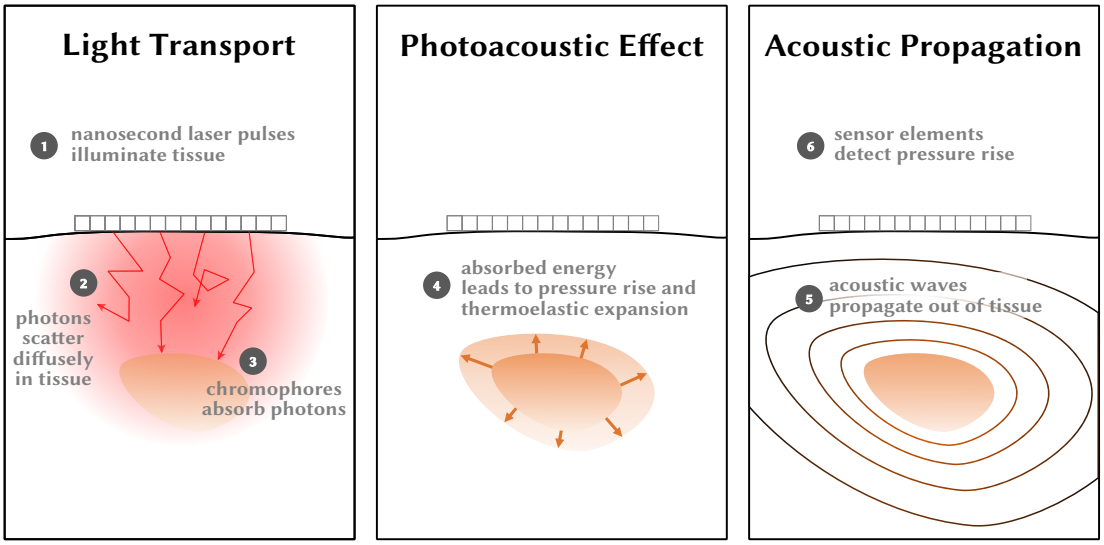


Figure 2.1: Principles of photoacoustic imaging: Light transport (1-3) is described by the radiative transfer equation (Sec. 2.1.1). The photoacoustic effect (4) and acoustic propagation (5 & 6) follow below. Figure credits to Tom Rix inspired by [18].

initially in the context of wireless sound transmission, active research into PA (or optoacoustic) imaging (Fig. 2.1) did not begin until the development of ultrasound transducers, pulsed lasers, and modern computing technologies [11]. As described in the previous section, when light interacts with tissue, it will eventually be absorbed. A portion of the absorbed energy converts into heat, which induces a localized pressure rise via thermoelastic expansion. This, in turn, generates an ultrasound wave that can be detected by an ultrasound transducer.

To generate a spatially resolved image, the detected signal (time series data) must be backprojected or reconstructed. By acquiring images at multiple wavelengths, characteristic patterns emerge that depend on the specific absorbers (i.e., molecular chromophores) present within the tissue. The concentrations of these chromophores can be estimated by analyzing the measured PA spectra, a process referred to as spectral unmixing.

Following this brief summary, the subsequent sections delve into the physical principles underlying PA image formation, primarily following the work of Wu and Wang [11] and Xia et al. [19].

Initial pressure in PAI To trigger the PA effect, the heating must be time-variant. This is typically achieved using a pulsed laser, although, in principle, continuous-wave lasers with modulated intensity can also be used. The fractional volume expansion dV/V caused by laser excitation is defined as (Eq. 12.3 in [11]):

$$\frac{d\mathbf{V}}{\mathbf{V}} = -\kappa \cdot p + \beta \cdot T. \tag{2.5}$$

Here, κ is the isothermal compressibility (Eq. 2.6), β is the thermal coefficient of volume expansion, and p and T are the changes in pressure and temperature, respectively. An overview of typical values for these variables is provided in Table 2.1. The compressibility κ can be further expressed in terms of the mass density ρ, the heat capacities at constant volume C_v and constant pressure C_p, and the speed of sound v_s:

$$\kappa = \frac{C_p}{\rho v_s^2 C_v}. \tag{2.6}$$

Under the conditions of thermal and stress confinement, Equation 2.5 simplifies to:

$$p_0 = \frac{\beta T}{\kappa}. \tag{2.7}$$

Thermal confinement is fulfilled if the laser pulse duration τ is shorter than the thermal relaxation time τ_{th} (Eq. 2.8), which implies that heat conduction is negligible:

$$\tau_{\text{th}} = \frac{d_c^2}{\alpha_{\text{th}}}. \tag{2.8}$$

Here, d_c represents the characteristic dimension of the absorbing structure, or as stated in [11], "the decay constant of the optical energy deposition, whichever is smaller," and α_{th} denotes the thermal diffusivity.

Analogously, stress confinement applies when the pulse duration is less than the stress relaxation time τ_s (Eq. 2.9), indicating negligible stress propagation:

$$\tau_s = \frac{d_c}{v_s}. \tag{2.9}$$

To proceed, the temperature rise T in Eq. 2.7 can be substituted using the absorbed optical energy density $A_e = \mu_a \cdot \Phi$ and the fraction of that energy converted to heat, η_{th}:

$$T = \frac{\eta_{th} A_e}{\rho C_v}. \tag{2.10}$$

Substituting Eq. 2.10 into Eq. 2.7 yields the final expression for the initial pressure p_0, which constitutes the initial pressure. It appears as a product of the Grüneisen parameter Γ, the heat conversion efficiency η_{th}, the optical absorption coefficient μ_a, and the local optical fluence Φ (Eq. 12.10 in [11]):

$$p_0 = \underbrace{\frac{\beta}{\kappa \rho C_v}}_{:=\Gamma} \eta_{\text{th}} A_e = \Gamma \, \eta_{\text{th}} \mu_a \Phi. \tag{2.11}$$

	Description	Typical value	
κ	isothermal compressibility	$\sim 5 \times 10^{-10} Pa^{-1}$	water & ST
β	thermal coef. of volume expansion	$\sim 4 \times 10^{-10} K^{-1}$	muscle
C_p	heat capacity at const. pressure	$\sim 4000\ J/(kgK)$	muscle
C_v	heat capacity at const. volume	$\sim 4000\ J/(kgK)$	muscle
ρ	mass density	$\sim 1000\ kg/m^3$	water & ST
v_s	speed of sound	$\sim 1480\ m/s$	water

Table 2.1: Typical values of tissue parameters involved in photoacoustic image generation [11]. ST: soft tissue

Acoustic wave propagation Having established how optical absorption leads to localized pressure generation via thermoelastic expansion, the subsequent propagation of the resulting initial pressure rise is now considered. Assuming a lossless acoustic medium, the resulting PA wave evolves according to the framework of linear acoustics, which include momentum conservation, mass conservation, and a pressure–density relationship:

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} &= -\frac{1}{\rho_0} \nabla p \\ \frac{\partial \rho}{\partial t} &= -\rho_0 \nabla \cdot \mathbf{u} \\ p &= v_s^2 \rho. \end{aligned} \tag{2.12}$$

Here, $\mathbf{u}$ denotes the particle velocity, ρ_0 is the ambient mass density, ρ is the acoustic density, and p represents the acoustic pressure.

Starting from the initial pressure distribution p_0 (Eq. 2.11) and assuming $\partial p_0 / \partial t = 0$ as outlined in [20], a second-order photoacoustic wave equation can be derived. This

formulation, as presented by Cox et al. 2005 [21], is given by:

$$\nabla^2 p - \frac{1}{v_s^2}\frac{\partial^2 p}{\partial t^2} = -\frac{\beta}{C_p}\frac{\partial H}{\partial t}. \tag{2.13}$$

In this expression, v_s is the speed of sound, β is the thermal coefficient of volume expansion, C_v is the specific heat capacity at constant volume, and H denotes the absorbed thermal energy per unit time and volume.

PAI image reconstruction With the PA wave equation and initial pressure distribution defined, the key question remains: how can a spatially resolved image be derived from the raw time-series data recorded by the detectors?

Over the years, numerous reconstruction algorithms have been proposed to address this task. Among the most prominent categories are:

- Universal back-projection [22] or delay-and-sum (DAS) [23, 24]
- Fast Fourier transform-based reconstruction [25, 26]
- Time reversal [27]
- Deep learning-based reconstruction algorithms [28]

To illustrate the fundamental idea behind PA image reconstruction, this paragraph provides a brief description of the DAS method, based on the work of Holfort et al. [23]. A full mathematical derivation of universal back-projection using Dirichlet Green's functions can be found in [22]. Additional important pre- and post-processing steps are outlined in Part II Section 5.1.3.3.

The objective of an ideal PA reconstruction algorithm is to reverse the acoustic wave propagation and perfectly recover the initial pressure distribution generated by the PA effect. However, this ideal is unattainable in practice due to several limitations: heterogeneous and often unknown tissue properties such as speed of sound, density, and acoustic attenuation; as well as limitations in detector geometry, angular sensitivity, and bandwidth.

In DAS reconstruction, which also referred to as beamforming, the received PA signals are delayed, weighted (apodized), and summed. The apodization weights per transducer element w are fixed and predetermined. To reconstruct the signal at a specific point $\vec{r}_p$ in the image plane, the signals from each detector element $m = 0, 1, \ldots, M-1$ are summed at the appropriate delay τ, which is determined by the distance between

the detector element $\vec{r}_m$ and the reconstruction point $\vec{r}_p$, as well as the speed of sound v_s:

$$\tau_m(\vec{r}_p) = \frac{\|\vec{r}_m - \vec{r}_p\|}{v_s}. \tag{2.14}$$

This leads to the beamformed signal $b(\vec{r}_p)$, with $s_m(t)$ representing the received signal at detector m:

$$b(\vec{r}_p) = \sum_{m=0}^{M-1} w_m s_m(\tau_m(\vec{r}_p)). \tag{2.15}$$

Spectral unmixing While reconstruction yields a spatial map of initial pressure values, further insight can be gained by analyzing multispectral data. If PA images are acquired at multiple laser wavelengths, a spectrum can be constructed for each pixel. In practice, such raw spectra are often difficult to interpret directly and typically lack intuitive clinical meaning. Nevertheless, PAI holds the promise of providing clinically relevant information about chromophores such as oxygenated hemoglobin (HbO_2), deoxygenated hemoglobin (Hb), and water, among others. According to Eq. 2.11, the measured spectrum correlates with the local optical absorption properties of the tissue.

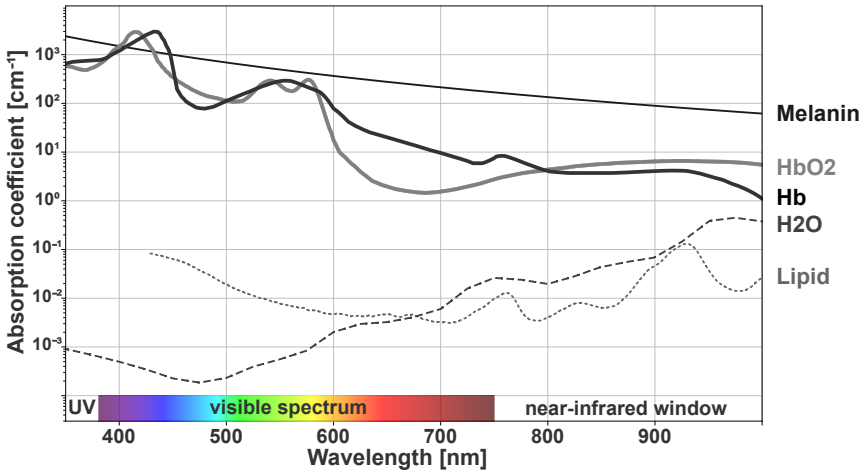


Figure 2.2: Absorption spectra of key biological chromophores in the visible and near-infrared wavelength range. The plot includes deoxygenated hemoglobin (Hb), oxygenated hemoglobin (HbO_2), water (H_2O), lipid, and melanin. The background color bar indicates the transition from ultraviolet (UV) through the visible spectrum to the near-infrared window. Figure credits to Tom Rix. Absorption spectra data adapted from [29].

Thus, the measured PA pixel value S at a given location x and wavelength λ is proportional to the absorption coefficient μ_a, as shown in [30]:

$$S(x,\lambda) \propto \mu_a(x,\lambda) = \sum_{n=1}^{N} c_i(x) \cdot \mu_{a,i}(\lambda). \tag{2.16}$$

Here, c_i represents the concentration of chromophore i, and $\mu_{a,i}(\lambda)$ its known absorption spectrum (Fig. 2.2). Assuming that the absorption coefficients are known and that all N absorbing chromophores within the tissue sample are identified, the individual concentrations c_i can be estimated by solving a linear system of equations. The system consists of the measured multispectral pixel vector $\hat{S}$ and a matrix A containing the absorption coefficients of all chromophores across the measured wavelengths. This approach is referred to as linear spectral unmixing [31].

$$\hat{S} = A \cdot c. \tag{2.17}$$

In the case of estimating concentrations of HbO_2 and Hb, the local blood oxygenation sO_2 can then be computed as:

$$sO_2 = \frac{c_{HbO_2}}{c_{HbO_2} + c_{Hb}}. \tag{2.18}$$

In practice, the number of measured wavelengths often exceeds the number of chromophores for unmixing. This makes the system overdetermined, and the the matrix A is no longer square, and its inverse does not exist. Instead, the Moore-Penrose pseudoinverse A^+ is used [32], which can be efficiently computed using methods such as singular value decomposition (SVD); a detailed derivation is available in [33].

Algorithm 2.1 Linear spectral unmixing (vectorized)

```
for all chromophores do
    for all wavelengths λ do
        Retrieve absorption value a_λ from spectral chromophore library
        A[λ][chromophore] ← a_λ
    end for
end for
Reshape multispectral image data into matrix B: shape = [wavelengths, pixels]
Compute pseudo-inverse A^+ of absorption matrix A
Compute concentrations: X ← A^+ · B          ▷ X = chromophore concentrations
```

In practical implementations, linear spectral unmixing is typically performed by

calculating the pseudoinverse via SVD (as implemented, for example, in SciPy), or alternatively by solving the least-squares optimization problem $\min_{\mathbf{x}} |A\mathbf{x} - \mathbf{S}|_2$ [34, 35] (Alg. 2.1).

For further reading and a thorough analysis of different linear unmixing algorithms, the reader is referred to my Bachelor thesis [36]. While linear spectral unmixing has proven effective in detecting substantial changes in oxygenation at superficial depths [37], its reliability diminishes rapidly for deeper tissue structures (Fig. 2.3). In general, the failure of linear unmixing is attributed to neglecting the contribution of light fluence (Eq. 2.11), which introduces wavelength-dependent attenuation and shadowing effects, a phenomenon known as spectral coloring [38].

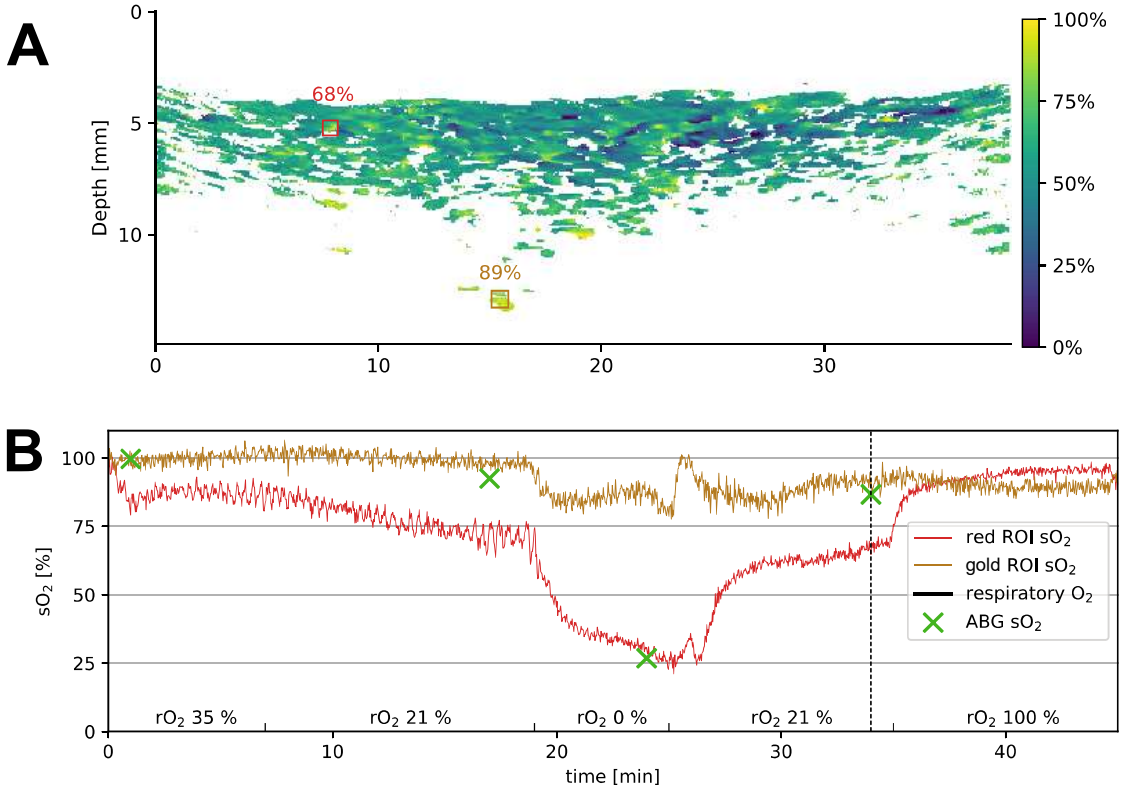


Figure 2.3: **A**: Oxygenation (sO_2) map of gyrencephalic brain, derived with linear spectral unmixing from a photoacoustic measurement [37]. **B**: Respiratory oxygen (rO_2) is varied over time (x-axis), which influences the measured sO_2 in the shallow red, and deep gold region of interest (ROI) in **A**. The dashed black line in **B** indicates the timepoint of the 2D photoacoustic image in **A**. For reference, arterial blood gas (ABG) [39] measurements are shown in green. Figure reused and adapted from my Bachelor thesis [36].

To overcome these limitations, current research in quantitative PAI explores explicit modeling of light fluence or the use of deep learning approaches for robust unmixing [40]. Despite these advances, true quantitative chromophore estimation in PAI remains an open challenge in the field (Barrier 7 in [6]).

Imaging device Various PAI systems exist, including microscopic, raster-scan mesoscopic, and tomographic modalities, each available as both commercial and custom-built devices. In this work, a commercial multispectral optoacoustic tomography (MSOT) system was employed. The MSOT Acuity Echo (iThera Medical GmbH, Munich, Germany) system specifications described herein are based on official datasheets, user training, and acoustic simulations of the transducer provided by iThera Medical, as well as detailed descriptions in [41].

The MSOT device employs a pulsed Nd:YAG laser tunable from 660 nm to 1300 nm with sub-10 ms tuning times and 1 nm spectral steps. The system's pulse repetition rate enables a PA frame rate of 25 Hz, supporting sequential multi-wavelength acquisition within a single scan. Data acquisition can be performed either as a snapshot capturing 8 frames per wavelength or as a continuous scan that measures one selected wavelength over time. The integrated US transducer comprises 256 piezoelectric elements configured in a linear 1D arc array with a 4 cm radius and a 125° angular aperture. The system samples at 40 MHz, with a center frequency of 4 MHz and a 52% bandwidth (6dB) [41]. Notably, the transducer elements are coated with a reflective layer to redirect backscattered laser light into the tissue, enhancing signal strength at greater depths.

The MSOT system allows for the acquisition of co-registered US and PA images. US images are available at an approximate frame rate of 5 Hz. Representative output from the imaging system is shown in Figure 7.1. The in-plane resolution ranges from 172 μm at the acoustic focus to approximately 200–350 μm toward lower depths and at the periphery of the field of view (FOV).

2.1.4 Model-based photoacoustic simulation

Simulating PA images typically involves three major steps: (1) tissue modeling comprising volume generation, (2) the assignment of optical and acoustic properties, and (3) device modeling, which entails simulating the image formation as an optical and acoustic forward process.

Tissue modeling refers to the creation of a digital volume with spatially varying optical and acoustic parameters, designed to emulate various tissue types (e.g., organs)

and physiological states (such as differing levels of oxygenation). This step is critical for generating realistic and physically meaningful simulations. For further reading on this topic, I refer to the recently published monograph by my former colleague Melanie Schellenberg [42], which offers a comprehensive treatment of tissue generation strategies for PA simulations. Additionally, Part II Section 7 provides a concrete example of manual, semantic segmentation-based tissue modeling as used within this work.

Optical simulation As described in the previous section, the analytical model governing photon transport in tissue is the RTE (Eq. 2.2). In practice, this equation is most commonly approximated using Monte Carlo simulation techniques, as detailed in [11].

Monte Carlo modeling approximates light transport by simulating the random walk of individual photons through a scattering and absorbing medium. Specifically, this involves (1) tracking each photon as it undergoes scattering events, (2) recording the deposited energy along its path, often allowing partial energy deposition at each step to improve computational efficiency, and (3) terminating the photon once a predefined condition (e.g., energy threshold or exit from the volume) is met.

The validity of this approach relies on the assumption that the physical quantity of interest can be accurately approximated by the expectation value of the underlying stochastic process [11, 43]. As such, Monte Carlo simulations remain the gold standard for modeling photon transport in highly scattering biological media.

Acoustic simulation To model the acoustic forward process in PAI, it is necessary to solve the photoacoustic wave equation (Eq. 2.13) or, alternatively, the corresponding system of first-order linear acoustic equations (Eq. 2.12), as outlined in the previous section and discussed in more detail by Treeby and Cox [20], and Cox et al. [21].

The latter approach is implemented in the widely used k-Wave toolbox [20], which employs the k-space pseudospectral method [44] in combination with fast Fourier transforms (FFTs) to compute spatial gradients. This method offers significant advantages in terms of memory usage and computational efficiency, making it suitable for simulating acoustic wave propagation in large, heterogeneous volumes [45].

Noise modeling An important step toward achieving more realistic PA simulations is the incorporation of noise modeling, which can, in principle, be applied at various stages of the simulation pipeline. It aims to capture signal variations introduced by

sources such as thermal or electronic noise [46, 47]. Mathematical formulations for modeling both optical and acoustic noise in PAI can be found in [48]. Several strategies have been proposed in the literature, ranging from adding Gaussian-distributed noise to approximate thermal noise, to including device-specific noise patterns obtained by water bath measurements [49, 50, 51].

2.2 Medical image registration

Since one of the key technical contributions of this work lies in registering (or stitching) 2D PAI slices into 3D volumes (slice-to-volume registration [52]), as well as enabling multimodal registration, the first part of this section is dedicated to the problem description, followed by the second part providing foundational principles, definitions, and evaluation measures of image registration. Finally, the third part addresses the algorithmic principles underlying 2D-to-3D ultrasound image registration.

Conceptually, the first part is structured based on the textbook *Numerical Methods for Image Registration* by Jan Modersitzki [53], which synthesizes fundamental work in image registration, including that of Maurer and Fitzpatrick [54], and builds upon research efforts from the 1980s and 1990s, particularly in medical imaging applications.

In essence, the image registration problem is defined as: "find an optimal geometric transformation between corresponding image data" [53]. Such scenarios commonly arise when image data is acquired at different times, from different devices, or from varying perspectives. In medical contexts, this includes applications such as pre- and post-operative comparisons, multimodal data fusion, and alignment of consecutive tissue sections.

To remain within the scope of this thesis, the focus is restricted to rigid registration involving translations and rotations, while non-rigid registration (nonlinear transformations and tissue deformation) is excluded.

2.2.1 Problem description

Mathematically, image registration is formulated as the task of minimizing a distance measure D between a fixed reference image F and a moving image M. To align the two images, a transformation φ and an intensity adjustment function g are applied. The function g accounts for differences in image intensity values, which commonly arise when F and M originate from different imaging modalities. In the case of rigid registration, the transformation φ consists of translation and rotation operations.

This leads to the following formal definition of the registration problem, as given in Problem 3.1 of [53] with Img(d) being the set of all images with spatial dimension d:

Problem

Given a distance measure D: Img(d)$^2 \rightarrow \mathbb{R}$ and two images F, M $\in$ Img(d), find a mapping $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and a mapping g : $\mathbb{R} \rightarrow \mathbb{R}$ such that

$$D(F, g \circ M \circ \varphi) = \min . \tag{2.19}$$

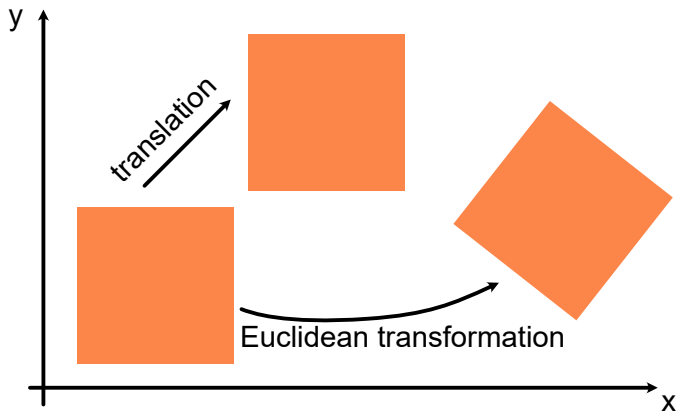


Figure 2.4: Example of rigid transformations in 2D space. The upper shape undergoes a pure translation, while the bottom shape is transformed using an Euclidean transformation i.e., a combination of translation and rotation. Figure inspired by [55].

First, we define the transformation mapping φ (Eq. 2.20) that aligns the moving image M to the fixed reference image F. Since this work is limited to rigid registration, φ is constrained to consist solely of translations and rotations (Fig. 2.4), excluding any deformations or shearing.

$$\varphi(x) = Qx + b \tag{2.20}$$

With $Q \in \mathbb{R}^{d \times d}$ denoting an orthogonal matrix with $\det(Q) = 1$, and $b \in \mathbb{R}^d$ as a d-dimensional translation vector. Determining the optimal transformation φ is often inherently an ill-posed problem, particularly in the absence of additional constraints or prior knowledge. The solution space may be large or underdetermined, especially when image information is sparse or noisy.

In the following paragraphs, the focus is placed exclusively on rigid, landmark-based registration, where discrete anatomical or artificial features are used to drive the alignment. For further details on other regularization strategies and non-rigid

extensions, the reader is referred to [53].

2.2.2 Landmark-based registration

Landmark-based registration simplifies the general image registration problem by reducing it to the alignment of a finite set of image features, referred to as landmarks, that must be mapped between the fixed and moving images. Landmarks are distinctive features that can be localized in the image with high accuracy. They may be soft (anatomical) landmarks identified by medical experts, or hard landmarks such as artificial fiducials. The registration task based on landmarks can be formally expressed as follows (Problem 4.1 in [53]):

Problem

Let $m \in \mathbb{N}$ and let the features $\mathscr{F}(\mathrm{F}, j)$ and $\mathscr{F}(\mathrm{M}, j)$, for $j = 1, \ldots, m$, be given. Find a transformation

$$\varphi \colon \mathbb{R}^d \to \mathbb{R}^d$$

such that

$$\mathscr{F}(\mathrm{F}, j) = \varphi(\mathscr{F}(\mathrm{M}, j)) \quad \text{for } j = 1, \ldots, m. \tag{2.21}$$

This transformation can be obtained by minimizing a distance measure D, as introduced in the general formulation of image registration, now defined as the sum of the norms between corresponding landmarks in the fixed image F and the transformed moving image $\varphi \circ \mathrm{M}$.

Registration errors When artificial fiducials are used to perform the registration, the quality of alignment is commonly quantified using the (weighted) root mean squared error (RMSE) between corresponding fiducials. This metric is known as the fiducial registration error (FRE) [56]:

$$\mathrm{FRE} = \sqrt{\frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \|y_i - \hat{y}_i\|^2}. \tag{2.22}$$

Here, w_i are optional weighting factors, y_i represents the fiducials in the fixed image, and $\hat{y}_i$ the corresponding fiducials in the transformed moving image. The number of fiducials is denoted by n. Misalignments after registration are primarily attributed to the fiducial localization error (FLE) [57], which describes inaccuracies in locating the true position of a fiducial point, often limited by the resolution of the imaging system.

Assuming that FLE is normally distributed, independent, and isotropic, minimizing the squared distances leads to a maximum likelihood estimate for the transformation, with $w_i = 1$ for uniform FLE and variable weights if FLE differs across fiducials [57]. Artificial landmarks are particularly well-suited for such methods, as they can be designed to produce high contrast, unambiguous signals and are inherently non-deformable.

However, clinical applications typically require accurate registration of the underlying target structures, such as organs, tumors, or bones, rather than fiducials. In such cases, anatomical landmarks must be used to quantitatively asses the registration, and their extraction is generally more complex. The RMSE computed between corresponding anatomical landmarks in the fixed and moving target region is referred to as the target registration error (TRE). Importantly, FRE and TRE are not necessarily correlated [56]; in fact, "deviations of FRE from its mean are uncorrelated to deviations of TRE from its mean" [57]. Nevertheless, unusually high FRE values can serve as indicators of potential issues in the registration setup.

Volume stitching One prominent application of landmark-based registration is volume stitching. The goal of volume stitching is to create seamless composite images by aligning multiple partially overlapping acquisitions. This process has numerous applications in computer vision and medical imaging, including video stabilization, panoramic image generation, and multimodal data fusion [55].

Volume stitching, in general, extends beyond pure landmark-based registration and may involve additional correction steps such as intensity normalization and image blending. For a comprehensive overview, see [55]. Broadly, there are two main categories of alignment strategies: (i) direct, pixel-based methods and (ii) feature-based methods. While direct methods offer greater precision for fine-scale alignment, feature-based approaches are more robust to larger-scale displacements and scene variation.

Recent developments in this field have increasingly focused on deep learning-based methods. In particular, unsupervised learning techniques show strong potential for more accurate and efficient stitching capabilities [58].

2.2.3 Principles of 2D to 3D image reconstruction

Having established how to perform registration between two image domains, the next question is how to utilize this positional information to reconstruct a coherent 3D

volume from a set of 2D slices. This part of the chapter is primarily based on the comprehensive review by Solberg et al. (2007) on 3D ultrasound reconstruction techniques [59]. For a broader overview of more recent technical developments, including tracking-based implementations for 2D-to-3D registration, the reader is referred to [60] and [61]. The translation of these methods into the PAI domain is discussed in the technical state-of-the-art section (Sec. 3).

Assuming that the spatial relationship between a set of 2D tomographic slices is known, e.g., via an external or image-based tracking method, the core challenge becomes reconstructing a consistent and anatomically meaningful 3D volume. A variety of approaches exist, and the appropriate method is often determined by the specific application and imaging goals. These include, but are not limited to, surface-based reconstruction, graph-based methods, and point cloud fusion.

For the scope of this thesis, the focus is limited to three categories of methods, as classified by Solberg et al. [59] and Rohling et al. [62], based on implementation characteristics rather than operational theory: (1) voxel-based methods, (2) pixel-based methods, as employed in this work for the optical pattern approach, and (3) function-based methods. These approaches provide the necessary framework for transforming spatially referenced 2D image data into 3D reconstructions and are particularly relevant for freehand imaging setups, such as those encountered in both ultrasound and PAI.

Voxel-based methods Voxel-based methods are voxel-centric, meaning that each voxel in the target 3D volume is independently assigned a value based on the surrounding 2D image data. The core idea is to evaluate every voxel location and determine its corresponding intensity value by referencing one or more pixels from the nearest 2D image slices. This assignment can be performed in two principal ways. First, through direct assignment, where a voxel is simply assigned the value of the nearest pixel. A representative example of this is the voxel nearest neighbor approach [63]. Second, through interpolation schemes, where voxel values are derived from a weighted combination of several neighboring pixels. A common strategy is to interpolate between the two closest image slices [64], where each voxel is assigned a weighted sum of the four closest pixels (from both slices) based on their distance to the voxel. The procedure for this interpolation-based voxel assignment is outlined in Algorithm 2.2, adapted from the appendix of [59].

Voxel-based methods offer the advantage of producing fully filled 3D volumes and are inherently well-suited for parallelization. However, their primary drawback lies in computational cost, as evaluating each voxel requires searching through potentially

Algorithm 2.2 Voxel-based mapping

for voxel v in 3D volume **do**
  Find nearest surrounding 2D slices S_1 and S_2
  for $i = 1$ to 2 **do**
    Compute the normal vector n_i from voxel v to slice S_i
    Compute the intersection point p_i on S_i using n_i
    Identify the four surrounding pixels of p_i on S_i
    Interpolate the intensity value I_i at p_i using bilinear interpolation
  end for
  Compute distances d_1 and d_2 from voxel v to slices S_1 and S_2
  Calculate final voxel intensity:

$$I_v = \frac{d_2 \cdot I_1 + d_1 \cdot I_2}{d_1 + d_2}$$

  Assign I_v to voxel v
end for

large sets of 2D slices, which can become expensive in terms of memory and runtime.

Pixel-based methods Pixel-based methods adopt a pixel-centric approach, whereby each pixel from a 2D image slice is assigned to one or multiple nearest voxels in the 3D reconstruction grid. These methods typically follow a two-step process.

The first step is the distribution step, in which each pixel is projected into 3D space and stored in the corresponding voxel(s), often with an associated weight. Pseudocode for this process is provided in Algorithm 2.3, adapted from the appendix of [59].

Algorithm 2.3 Pixel-based mapping

for each image slice I **do**
  for each pixel p in I **do**
    Compute the 3D position P of pixel p in the global volume coordinate system
    Find the nearest voxel v to P
    Add pixel intensity $I(p)$ to voxel sum buffer: $v.\text{sum} \mathrel{+}= I(p)$
    Increment voxel counter: $v.\text{count} \mathrel{+}= 1$
  end for
end for
for each voxel v in the volume **do**
  if $v.\text{count} > 0$ **then**
    Compute average intensity: $v.\text{value} = \dfrac{v.\text{sum}}{v.\text{count}}$
  end if
end for

The second step is hole-filling, which addresses voxels left unassigned during distribution. These empty voxels are filled based on surrounding voxel values, either by interpolating from the local neighborhood or by applying 3D smoothing kernels centered around filled voxels (Fig. 2.5). Hole-filling commonly incorporates a maximum fill distance to limit extrapolation, which may result in gaps remaining within the reconstructed 3D volume. Alternatively, distribution and hole-filling can be unified into a single step by distributing each pixel's value into a local neighborhood of voxels using, for example, spherical kernels [65].

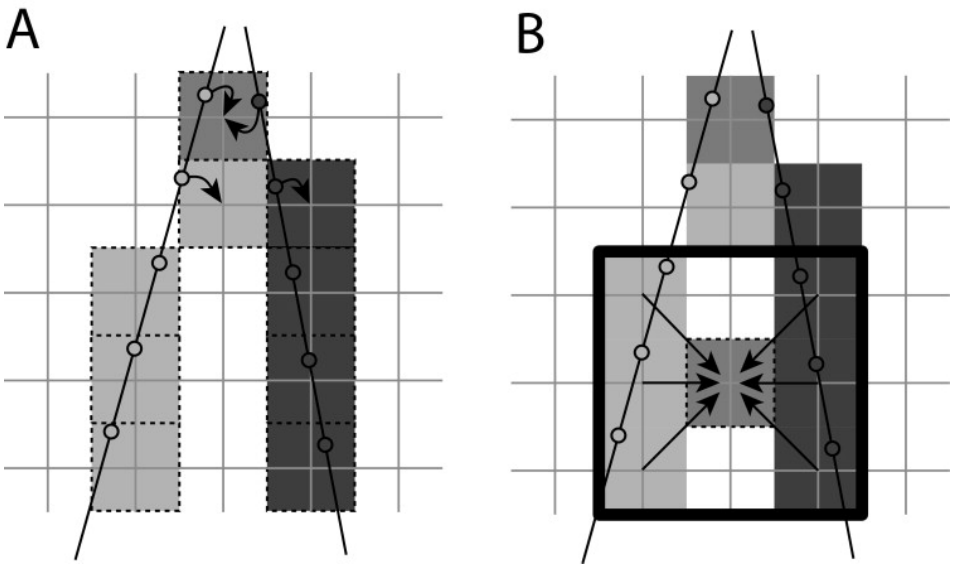


Figure 2.5: "[Pixel-Based method distribution step] PBM DS: [Pixel-nearest neighbor] PNN bin-filling stage, each voxel is assigned the value of the intersecting pixel(s). This is done by traversing each pixel and assigning the pixel value to the voxel occupying the pixel position. Multiple contributions to the same voxel are handled either by using an average of all values or by storing the maximum or the latest value. (b)[Pixel-based method hole filling step] PBM HFS: [Pixel-nearest neighbor] PNN hole-filling from a neighborhood around the hole, each empty voxel is assigned a value based on the nearest already filled voxels. This value may either be an average or median, or an interpolation between the two closest nonzero voxels. (In both illustrations, the input images are illustrated by lines and the points in the lines illustrate the centers of the pixels. The 2D grids mark the centers of the voxles in the 3D voxel grids.)" [59] Reprinted from Ultrasound in Medicine & Biology, Vol 33.7, Solberg et al., Freehand 3D Ultrasound Reconstruction Algorithms—A Review, Pages 991-1009, 2017, with permission from Elsevier.

Pixel-based methods are intuitive and computationally efficient, particularly well-

suited for sparse or irregular scanning trajectories. However, due to their reliance on discrete pixel distribution, they may leave gaps in the final volume, especially in regions with limited scan coverage.

Function based methods Function-based methods take a continuous approach to 3D reconstruction by modeling the intensity values of input pixels as a continuous function defined over the target volume. These functions are subsequently evaluated at regular voxel grid intervals, assigning an intensity value to each voxel in the reconstructed volume.

A commonly used strategy is radial basis function interpolation, which approximates subsegments of the volume using spline functions [62]. This approach offers smooth interpolation across spatial regions and is particularly effective when dealing with irregularly spaced input data. Pseudocode for this method is shown in Algorithm 2.4, adapted from the appendix of [59].

Algorithm 2.4 Function-based mapping

```
1: Divide the voxel array into rectangular segments
2: for each segment S do
3:     Define a window W around S
4:     for each direction d in {x, y, z} do
5:         Grow W in direction d until it includes sufficient neighboring data points
6:     end for
7:     if S contains too many data points then
8:         Subdivide S into smaller segments
9:     end if
10:    Collect all data points within window W
11:    Fit a Radial Basis Function RBF to the data in W
12:    Evaluate RBF at regular voxel intervals within S
13:    Assign computed values to corresponding voxels
14: end for
```

Function-based methods are especially valuable when smooth and continuous volume representations are required. However, they may be computationally more intensive compared to voxel- or pixel-based approaches.

Nyquist-Shannon sampling theorem The Nyquist–Shannon sampling theorem provides the fundamental criterion for how densely 3D space must be sampled in order to accurately reconstruct a volume from a set of 2D images [66]. This section is based

on the work of Por et al. [67], which contextualizes the theorem within the scope of medical imaging.

For a bandlimited signal, i.e., a signal whose frequency content does not exceed a finite maximum frequency f_{max}, the Nyquist–Shannon theorem states that the original signal can be perfectly reconstructed only if it is sampled at a rate f_{sample} that is at least twice the maximum frequency:

$$|\hat{X}(f)| = 0 \quad \forall\, |f| > f_{\max} \tag{2.23}$$

$$f_{\text{sample}} \geq 2 f_{\max}. \tag{2.24}$$

Here, $\hat{X}(f) = \mathcal{F}X(t)$ denotes the Fourier transform of the bandlimited signal X, as described in [67].

While the theorem is commonly formulated in the frequency domain, it is equally applicable in the spatial domain. In imaging, the sampling process corresponds to discretizing continuous intensity values into pixels. Applying the theorem in this context implies that the pixel size must be at most half the desired spatial resolution to ensure all image details can be reconstructed without aliasing.

This principle underpins sampling requirements not only during image acquisition but also in 3D volume reconstruction from sequential 2D slices. Ensuring appropriate spatial sampling density is therefore crucial for accurate and artifact-free reconstruction.

2.3 Data science

Any algorithmic or technological advancement introduced in Part II, as well as the clinical data presented in Part III, must be evaluated, assessed, and quantified. This chapter serves to establish the statistical and analytical foundation necessary for that purpose. It begins with an overview of commonly used quantification metrics relevant to image analysis and prediction tasks, followed by background information on selected models and evaluation tasks applied throughout this thesis. In essence, this chapter provides the statistical backbone of the presented work.

2.3.1 Statistical evaluation

Evaluating performance of algorithms, as well as assessing and quantifying clinical results, relies heavily on the use of various evaluation metrics. Throughout the text, the terms measures and metrics are used synonymously to describe methods or formulas that quantify or assess results. Selecting appropriate metrics and performing statistical testing is a non-trivial task. The process is often complex and fraught with potential pitfalls that can lead to misinterpretation or overestimation of performance. For comprehensive guidance on metric selection and interpretation within the field of medical image classification, the reader is referred to the metric recommendations outlined in [68] and the discussion of common pitfalls presented in [69]. These two works form the foundation for the following paragraphs unless stated otherwise.

Statistical testing This paragraph provides a concise introduction to essential statistical terminology, based primarily on the article series by Ranganathan et al. [70, 71] and a textbook by Wassertheil-Smoller [72], which offers a particularly practical entry point into biostatistics for applied medical research.

The first fundamental concept is the *sample*. In statistics, a sample refers to a representative subset of the broader *population* (e.g., all humans) that is examined in a study. The goal is to draw conclusions about the population based on findings observed in the sample [73].

A central tool to quantify the certainty of such extrapolations is statistical testing, particularly in the form of hypothesis testing. Here, we attempt to find sufficient statistical evidence to reject the null hypothesis, the default assumption that there is no effect or no difference. Rejecting the null hypothesis lends support to our finding, though it never constitutes definitive proof [72]. For example, if we wish to demonstrate that an algorithm can accurately classify a disease, hypothesis testing enables us to

challenge the null hypothesis that classification is equivalent to random guessing. Similarly, in correlation analysis, such as with Pearson correlation, the null hypothesis posits that there is no linear relationship between two variables.

Before exploring these concrete statistical methods in detail, we first examine how to quantify the confidence in our findings.

Confidence intervals A confidence interval (CI) is a range of values, calculated from sample data, that is likely to contain the true population parameter with a specified level of confidence, usually 95% [74]. CI calculations assume random sampling from the population. In contrast to the widely reported p-value, which provides evidence against the null hypothesis, CIs allow "to compare magnitude of a difference" [74]. This distinction is crucial, as two distributions can be statistically significantly different while still lacking clinical relevance. Especially for small sample sizes, *Metrics Reloaded* recommends reporting confidence intervals for all metrics to capture uncertainty of the metric and improve interpretability.

A major tool for deriving CIs is bootstrapping [75], which, while relatively simple, is a powerful method for estimating confidence intervals. Bootstrapping follows the basic principles of a Monte Carlo method, as introduced in the photoacoustic optical forward simulation (Sec. 2.1.4). A bootstrap sample consists of a random resampling of the original data with replacement. The distribution of a statistic (e.g., the difference in means) across many such samples will approximate a normal distribution, as per the central limit theorem. Consequently, the 95% CI can be estimated as the 2.5% and 97.5% percentiles of the ranked bootstrap estimates. Later in this section, an additional method for deriving CIs will be presented using Fisher's z' transformation.

Pearson correlation Pearson correlation, also known as momentum correlation, denoted by r when applied to a sample, quantifies the strength and direction of a linear relationship between two continuous variables, X and Y [76]. The coefficient is computed by dividing the covariance of the two variables by the product of their standard deviations, resulting in a normalized value between -1 and 1 [77]:

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})\,(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2}}. \tag{2.25}$$

Here, n represents the number of paired observations, x_i and y_i are the individual scores for variables X and Y, and $\bar{x}$, $\bar{y}$ are their respective means. A value of $r = 1$

indicates perfect positive linear correlation, $r = -1$ indicates perfect negative linear correlation and $r = 0$ implies no linear relationship between the variables. Pearson correlation assumes that the relationship between the variables is linear and that the data are approximately normally distributed. It is also sensitive to outliers, which can disproportionately influence the computed correlation coefficient [78].

Spearman correlation In contrast to Pearson correlation, Spearman correlation is a nonparametric measure of association between two variables. It assesses the strength and direction of a monotonic relationship rather than a strictly linear one, making it less sensitive to outliers and deviations from normality [79]. Spearman correlation operates on the rank order of the data rather than the raw values. That is, it evaluates how well the relationship between two variables can be described by a monotonic function. Mathematically, Spearman correlation is equivalent to applying the Pearson correlation formula (Eq. 2.25) to the ranks of the two variables instead of their original values.

Fisher z' transformation To compute CIs for the Pearson correlation coefficient, the Fisher z' transformation is commonly employed (Eq. 2.26) [77]. This transformation stabilizes the variance of the sampling distribution of r, making it approximately normally distributed. Although originally developed for Pearson correlation, it may also be applied to Spearman correlation under certain conditions [80]. A key assumption underlying the use of Fisher's transformation is that the variables X and Y are jointly bivariate normally distributed. The following steps, based on the procedure described in [77], are used to derive the confidence interval for r:

(1) Apply the Fisher transformation to r:

$$z' = .5 * \ln\left(\frac{1+r}{1-r}\right). \tag{2.26}$$

(2) Compute the 95% confidence interval in the transformed z' domain:

$$z' \pm 1.96 * \sigma_{z'}, \tag{2.27}$$

using the approximate standard error $\sigma_{z'}$, defined as:

$$\sigma_{z'} = 1/\sqrt{N-3}. \tag{2.28}$$

(3) Finally, convert the resulting confidence interval back to the original correlation

scale using the inverse Fisher transformation:

$$r = \frac{\left(e^{2z'}\right) - 1}{(e^{2z'}) + 1} \tag{2.29}$$

True negative and true positive rate Turning back to assessing prediction results, a true positive (TP) prediction refers to a sample that is correctly predicted as positive and indeed belongs to the positive class. Similarly, a true negative (TN) is a sample correctly predicted as negative and truly belonging to the negative class.

In contrast, a false positive (FP) occurs when a sample from the negative class is incorrectly predicted as positive. Conversely, a false negative (FN) is a sample from the positive class that is incorrectly predicted as negative.

Two commonly used classification performance metrics derived from these definitions are specificity and sensitivity.

Specificity, also known as the true negative rate, measures the proportion of actual negative samples that are correctly identified as such [68]. It is defined as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{2.30}$$

Sensitivity, or the true positive rate, quantifies the proportion of actual positive samples correctly identified as positive [68]. It is defined as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2.31}$$

Together, these metrics provide insight into a classifier's ability to correctly distinguish between the positive and negative classes, which is particularly important in imbalanced datasets or clinically sensitive contexts.

Accuracy Accuracy is a metric ranging from 0 to 1 that quantifies the proportion of correctly classified samples among all samples [68]. It is defined as [68]:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{2.32}$$

Although accuracy is commonly used for biomedical classification tasks, it is prevalence-dependent and can be misleading in cases of class imbalance, where one class dominates the dataset [68]. In such scenarios, high accuracy may be achieved simply by predicting the majority class, without truly capturing the model's discriminative ability.

Area under the receiver operating characteristic curve The area under the receiver operating characteristic curve (AUROC) is a widely used metric for evaluating the separability of binary classes, such as distinguishing between healthy individuals and patients [68, 81]. As illustrated in Figure 2.6, AUROC quantifies how well a classifier can differentiate between positive and negative samples across all possible decision thresholds.

AUROC is a multi-threshold metric, meaning it evaluates performance across the full range of classification thresholds rather than relying on a single decision boundary. For this reason, it is considered the default recommendation for assessing performance in medical image-level classification tasks [68].

One of the key advantages of AUROC is its intuitive interpretability: a higher AUROC indicates better class separability, making it especially popular in clinical research. However, caution is advised when applying AUROC in settings with substantial class imbalance, as it may give overly optimistic performance estimates for minority class detection [68].

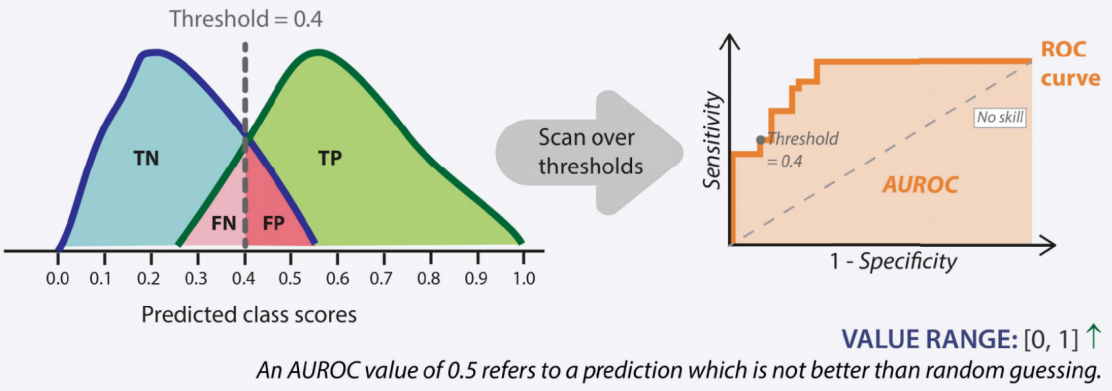


Figure 2.6: The area under the receiver operating characteristic curve (AUROC) is a multi-threshold metric that quantifies class separability. It evaluates a classifier's ability to distinguish between positive and negative classes across all possible decision thresholds. The curve is constructed by plotting the true positive rate against the false positive rate. Reproduced with permission from Springer Nature from Nature methods 21.2, Maier-Hein, et al., Metrics reloaded: recommendations for image analysis validation, 195-212, 2024 [68].

Positive likelihood ratio Likelihood ratios provide a diagnostic measure that is independent of disease prevalence, making them particularly valuable for clinical

decision-making [82]. The positive likelihood ratio (LR^+) quantifies how much more likely a positive test result is to occur in individuals with the condition compared to those without. As defined by [68], LR^+ represents "the factor by which a positive prediction occurs more frequently among actual positive samples than among actual negative samples," and is calculated as:

$$LR^+ = \frac{\text{Sensitivity}}{1 - \text{Specificity}}. \tag{2.33}$$

Here, sensitivity and specificity are defined in Equations 2.30 and 2.31, respectively. This measure is particularly useful when assessing the clinical utility of a classifier across diverse populations, regardless of underlying prevalence rates.

Dice similarity coefficient The Dice score, also known as the Dice similarity coefficient (DSC), is one of the most widely used overlap-based metrics for evaluating segmentation performance in medical imaging. It measures the degree of spatial overlap between a reference region A and a predicted segmentation B (Fig. 2.7), and is defined on a scale from 0 to 1, where 0 indicates no overlap and 1 indicates perfect agreement.

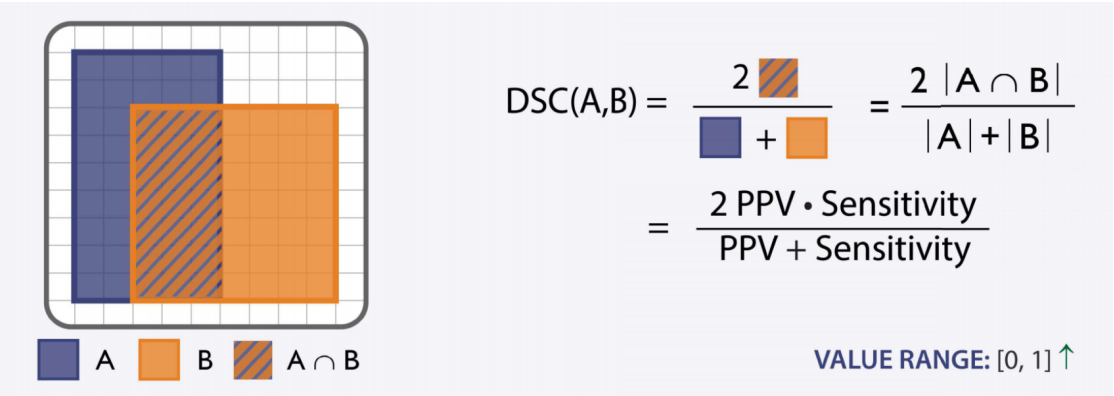


Figure 2.7: Illustration of the Dice similarity coefficient (DSC), an overlap-based metric used to evaluate segmentation accuracy. The DSC is defined for a reference region A and a predicted region B, and quantifies their spatial agreement. It can also be expressed in terms of the positive predictive value (PPV), $PPV = \frac{TP}{TP+FP}$, and sensitivity, where TP denotes true positives and FP false positives. Reproduced with permission from Springer Nature from Nature methods 21.2, Maier-Hein, et al., Metrics reloaded: recommendations for image analysis validation, 195-212, 2024 [68].

While DSC is highly popular due to its intuitive interpretation and straightforward

computation, it has limitations, particularly in scenarios involving small anatomical structures or significant variability in structure sizes. In such cases, the metric can become unstable or biased, especially under conditions of class imbalance [68].

Hausdorff distance The Hausdorff distance $H(A, B)$ is a boundary-based metric commonly used to evaluate segmentation performance by quantifying the worst-case spatial discrepancy between two sets. It is defined as the maximum distance from any point on the boundary of the predicted segmentation B to the closest point in the reference segmentation A, or more generally [83]:

$$H(A, B) = \max\left(h(A, B), h(B, A)\right). \tag{2.34}$$

Here, $A = \{a_1, ..., a_p\}$ and $B = \{b_1, ..., b_p\}$ are finite sets of points typically on the boundaries of the segmented objects, and $||\cdot||$ denotes the underlying norm used for distance computation:

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|. \tag{2.35}$$

Given its boundary-focused nature, [68] recommends using boundary-based metrics like the Hausdorff distance in conjunction with overlap-based metrics such as the DSC. Furthermore, using quantiles (e.g., the 95th percentile Hausdorff distance) instead of the maximum can provide a more robust and stable estimate in the presence of spatial outliers.

2.3.2 Modeling approaches

This section introduces fundamental modeling techniques that are essential for interpreting the clinical application results presented in Part III. The discussion begins with regression and linear mixed models, followed by an overview of clustering and classification methods.

Linear mixed models Throughout the previous section, it was explored how to assess whether two variables are linearly correlated using measures such as Pearson and Spearman correlation. Modeling linear relationships that also incorporate so-called *random effects* takes this a step further, which will be essential for the later analysis of clinical data.

This paragraph draws on foundational concepts introduced in the first chapter of the book *Linear and Generalized Linear Mixed Models and Their Applications* by Jiang and

Nguyen [84], providing a high-level overview of linear mixed modeling.

We begin with a standard linear regression model:

$$y = X\beta + \varepsilon. \tag{2.36}$$

- y: Vector of observed outcomes (response variable).
- X: Matrix containing the predictor variables (covariates).
- β: Vector of unknown fixed regression coefficients.
- ε: Vector of random errors or residuals.

However, in cases such as longitudinal clinical studies, multiple observations from the same individual over time are inherently correlated. To account for this, we assume that some coefficients in β are random. This leads to the formulation of the general linear mixed model (LMM) as:

$$y = X\beta + Z\alpha + \varepsilon. \tag{2.37}$$

In this extended formulation, β represents the *fixed effects* that capture population-level trends, associated with the known design matrix X. The vector α represents the *random effects*, which model group- or subject-specific deviations, linked to the known matrix Z. Both α and ε are unobservable, uncorrelated random variables, typically assumed to have mean zero and finite variance [84].

This structure enables hierarchical modeling, which is one of the main strengths of LMMs in clinical data science.

Dimensionality reduction In contrast to analyzing simple linear relationships between two variables, working with PAI data or high-dimensional data in general presents the challenge of limited intuitive understanding. Dimensionality reduction techniques such as principal component analysis (PCA) [85], uniform manifold approximation and projection for dimension reduction (UMAP), and t-distributed stochastic neighbor embedding (t-SNE) [86] provide a means to address this issue by projecting the data into a lower-dimensional space, making it more visually interpretable. The key objectives of PCA, as outlined by [85], are to (1) extract meaningful information, (2) compress the data, (3) simplify representation, and (4) analyze the underlying structure. PCA achieves this by finding a linear transformation from the original high-dimensional space to a new set of orthogonal axes known as principal components PC_i, each defined

as a weighted combination of the original variables Var_j with associated weights w_{ij} [87].

The first principal component captures the direction of maximum variance in the data, while the second principal component captures the direction of the next highest variance, subject to being orthogonal to the first. This transformation can be expressed as:

$$PC_i = \sum_{j=1}^{J} w_{i,j}\text{Var}_j, \quad \text{with } i = 1, \ldots, J. \tag{2.38}$$

Given a centered data matrix X (with columns representing variables and rows representing observations), the principal components can be derived using SVD on the covariance matrix $X^T X$ [33].

A common practical challenge in PCA is deciding how many principal components to retain. Two widely used strategies are: (i) setting a cumulative variance threshold and retaining the number of components needed to reach it, or (ii) using a Scree plot, in which the eigenvalues λ or the explained variance ratios (Eq. 2.39) are plotted against the component index. The optimal number of components can then be determined by identifying the *elbow* point in the plot (Fig. 2.8) [33].

$$\text{Explained variance ratio}_i = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j}. \tag{2.39}$$

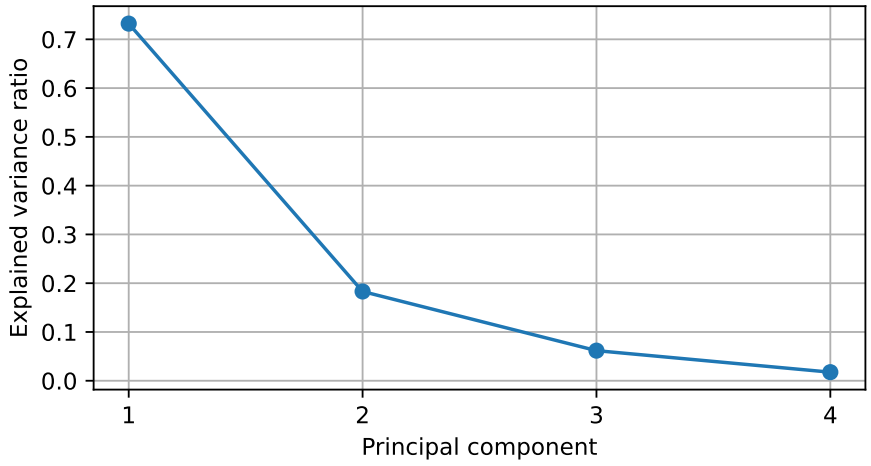


Figure 2.8: Scree plot to derive the number of principal components to select. Plotted is the explained variance ratio ($\sim$ Eigenvalue) against the principal component. An elbow after the second component is visible, so in this example, one would keep the first two and drop the 3rd and 4th components.

Random forest classification Classification refers to the task of assigning sample data to one of N distinct classes based on a decision rule or predictive algorithm [88]. While the previous sections focused on evaluating classification performance and selecting appropriate metrics, this section turns to the process of constructing decision rules, illustrated through the example of decision trees and their ensemble variant, random forest classification.

Originally proposed by Breiman [89], the random forest algorithm is applicable to both classification and regression tasks. At its core, it builds upon the concept of decision trees, which are binary trees that recursively split data based on feature-based decision rules. The process begins at a root node, which contains the entire dataset. At each subsequent internal node, the data is split into two child nodes based on a binary condition, continuing until a specified stopping criterion (such as maximum tree depth or minimum number of samples) is met. Each leaf node of the tree is then assigned a class label, typically based on the majority class of the training samples it contains. When classifying a new instance, the input is passed through the tree using the same decision rules as during training, eventually reaching a leaf node that yields the predicted label [90].

A common strategy for defining decision rules involves the use of an impurity function $I(t)$, which quantifies how mixed the classes are at a given node t. Splits are chosen to maximize the reduction in impurity, denoted as $\Delta I(t)$, calculated as follows:

$$\Delta I(t) = I(t) - a_R I(t) - a_I I(t). \tag{2.40}$$

Here, a_l and a_r represent the proportion of instances from the parent node t assigned to the left and right child nodes, respectively. Despite their interpretability and ease of use, decision trees are prone to high variance, since small changes in the training data can lead to substantially different tree structures [90]. To mitigate this instability, random forests aggregate the predictions of an ensemble of N independently trained decision trees. Each tree is typically trained on a bootstrapped subset of the data with random feature selection at each split. The final prediction is obtained through majority voting (for classification) or averaging (for regression) across all trees [91]. The topic of neural networks, including deep learning-based classifiers, is beyond the scope of this work. For a comprehensive overview of neural networks and deep learning-based classifiers, readers are referred to the book *Deep Learning* by Ian Goodfellow et al. [92].

2.4 Clinical background

The primary motivation behind the presented technical developments is their prospective clinical application. A comprehensive understanding of the relevant clinical context is essential to assess both the need and the potential impact of such innovations, including their limitations (Sec. 3.3.2). This chapter provides foundational insights into the clinical applications under investigation, specifically focusing on radiotherapy in the treatment of head and neck cancer taken from my first author publication [93], as well as peripheral artery disease (PAD) corresponding to the clinical contributions C1 and C2.

2.4.1 Head and Neck Cancer

"Squamous cell carcinoma represents the predominant subgroup of head and neck [...] tumor malignancies [94] with an annual occurrence of about 890,000 diagnoses representing about 4.5% of new cancer cases worldwide and resulting in around 450,000 fatalities annually (GLOBOCAN estimate) [95]. These tumors are characterized by large tumor volume and local invasion, usually with evidence of a high rate of metastasis to regional lymph nodes [96]. Besides surgery, primary radiochemotherapy is an equivalent treatment choice for locally advanced head and neck squamous cell cancer [...] with corresponding 5-year overall survival rates of around 50% [97] and locoregional control rates of approximately 60% at 2-years [98]. Individualized decision-making regarding initial therapy, sequencing, and administration requires expertise in balancing local tumor control, general patient morbidity, toxic effects of therapy, and functional preservation of treated structures.

Currently, there is no established routine of clinical evaluation during radiotherapy (RT) aimed at assessing early treatment success, particularly with regard to molecular parameters. The first treatment response evaluation typically occurs three months after RT and involves examination by an ear, nose, and throat [...] specialist, along with imaging (magnetic resonance imaging (MRI) or computed tomography (CT)) [99, 100]. Therefore patients receive a CT scan to evaluate tumor response following the RECIST 1.1 criteria [101], undergo a clinical examination, and report therapy-related side effects using the Common Terminology Criteria for Adverse Events [...] [102] questionnaire.

The main effects of RT occur on a molecular level, involving processes such as inflammation, revascularization, and reoxygenation. Molecular changes arise as a result of the interaction of ionizing radiation with tissues and cells, triggering a cascade of responses that may either resolve quickly (early molecular changes) or lead to long-term

modifications (late molecular effects). Early effects emerge within seconds to hours after radiation exposure and involve processes primarily related to deoxyribonucleic acid (DNA) damage and the activation of cellular stress responses, such as reactive oxygen species [...] production, leading to earlier clinical manifestations such as mucositis and radiodermatitis in the [head and neck] area [103]. Late molecular changes can occur weeks to years after radiation exposure and result from persistent damage, chronic oxidative stress, and long-term cellular dysfunction. These effects often reflect the tissue's inability to completely repair damage and adapt, for instance, endothelial dysfunction and vascular damage in which radiation-induced vascular injury triggers chronic inflammation and reduced perfusion. This, in the long-term, contributes to late fibrosis, organ dysfunction, or necrosis, as it is seen in patients after RT, who report persistent xerostomia. Various factors can influence the biological effects of radiotherapy in human tissues. For instance, the presence of oxygen enhances the radiotherapeutic effect by prolonging the lifespan of free radicals and reducing the cells' ability to repair sublethal DNA damage [104, 105]. Tumor tissue often exhibits intratumoral hypoxia, resulting in radioresistance, which can be addressed by fractionating radiotherapy to leverage reoxygenation effects. Hence, there's a significant need for patient selection, particularly in identifying those who may not benefit from conventional RT and should receive adapted or alternative treatment approaches like dose escalation or salvage surgery, or those who exhibit deep responses and could be candidates for RT de-escalation to minimize side effects. This approach aligns with the concept of "personalized medicine" as a fundamental principle in modern treatment planning [106].

One of the major side effects, salivary gland damage, stands as the prevailing long-term complication of both RT and chemoradiotherapy in [head and neck] cancer treatment. Xerostomia can result in notable symptoms like dry mouth and contribute to additional complications such as dental caries and nutritional challenges. Alterations in saliva quantity and composition shortly after starting radiation therapy indicate that these glands undergo acute and delayed responses [107]. There is no standard objective method to assess xerostomia [108] and we lack imaging modalities that enable us to directly assess early treatment success beyond tumor shrinkage" [93].

2.4.2 Peripheral artery disease

Peripheral artery disease (PAD) is one of the most prevalent cardiovascular conditions in the elderly population. It affects approximately 6% of individuals over the age of 25,

with prevalence increasing substantially with age reaching 13% in those aged 75–79 years [109]. As of 2015, PAD affected an estimated 237 million people worldwide, with smoking recognized as one of its major risk factors. Notably, only about 22% of PAD cases are symptomatic, highlighting the challenge of early clinical detection [110].

PAD is characterized by atherosclerotic occlusions in the arteries of the lower extremities. These vascular blockages can result in symptoms such as intermittent claudication, pain in the legs during walking, as well as at rest in more advanced cases [111, 112]. Chronic malperfusion due to long-standing arterial obstruction may also lead to muscular damage and impaired tissue function [113].

Early diagnostic tools for PAD include the ankle-brachial index (ABI), which compares blood pressure in the ankle and arm, and color-coded duplex sonography, both of which are non-invasive and widely used in clinical practice [112]. In cases where initial findings are inconclusive or interventional planning is required, invasive imaging modalities such as contrast-enhanced angiography are often employed to confirm the diagnosis [112].

Despite the availability of these diagnostic methods, diagnosing PAD remains challenging in patients with nonspecific or atypical leg pain during ambulation [112, 114]. The clinical classification of PAD was first formalized in 1952 by the European Society of Cardiovascular Surgery and is commonly referred to as the Fontaine classification [115]. It defines four major stages of disease progression:

I Asymptomatic

II Intermittent claudication

III Rest pain

IV Tissue necrosis or gangrene

Management of PAD depends on disease severity. For early-stage patients, first-line treatment includes lifestyle modification, particularly smoking cessation, and supervised exercise programs designed to improve pain-free walking distance [111]. In more advanced stages, patients with severe claudication or critical limb ischemia may require surgical revascularization procedures [111]. In extreme cases amputation may be necessary to prevent life-threatening complications [116].

3 Related work

This chapter provides an overview of the state of the art developments relevant to this thesis and is structured into a technical and clinical part. The technical state of the art is aligned with the main technical contributions of this work (T1–T3), beginning with advancements in 3D PAI (Sec. 3.1.1), followed by progress in multimodal image registration (Sec. 3.1.2), and concluding with developments in digital twin modeling (Sec. 3.1.3). The clinical state of the art is organized according to the primary clinical applications (C1 & C2), covering work related to monitoring tissue response to radiotherapy (Sec. 3.2.1) and detecting PAD (Sec. 3.2.2). Additionally current limitations of clinical PAI (Sec. 3.2.3) and a short conclusion (Sec. 3.3) are presented.

3.1 Technical state of the art

The majority of clinically used PA systems have integrated co-registered US, which has proven particularly valuable for real-time anatomical guidance during clinical procedures. As a result, many of the advancements in volumetric imaging have been developed in ways that are directly transferable to both PAI and US, with several methods either having been originally proposed for US or having demonstrated effectiveness specifically in US applications.

Therefore, the following chapter is not limited to techniques exclusively validated in PAI, but instead has incorporated key approaches from the ultrasound domain where relevant. However, the overall presentation has remained PAI-centric; for methods that have been initially introduced for US imaging, their potential or demonstrated applicability to PAI has been explicitly discussed.

3.1.1 Volumetric photoacoustic imaging

Fenster and Downey [117] have identified a major limitation of conventional 2D ultrasound imaging: its heavy reliance on the clinician's experience, particularly when mentally reconstructing a 3D tissue structure from 2D slices. As they have noted, this challenge arises "from using a spatially flexible 2-D imaging technique to view 3-D anatomy" [117]. In addition to this high operator dependency, low reproducibility associated with 2D imaging has further motivated the transition towards 3D imaging approaches, a trend that has also been recognized in the field of PAI [118].

3D PAI has become an active and evolving research area, as highlighted in a recent review by Lee et al. [119], which has provided a comprehensive overview of handheld

volumetric PAI techniques. Lee [119] proposed a categorization of handheld volumetric photoacoustic systems into four distinct types, with a strong focus on technical implementation (Fig. 3.1): Mirror scanning, direct scanning, mechanical scanning, and freehand scanning.

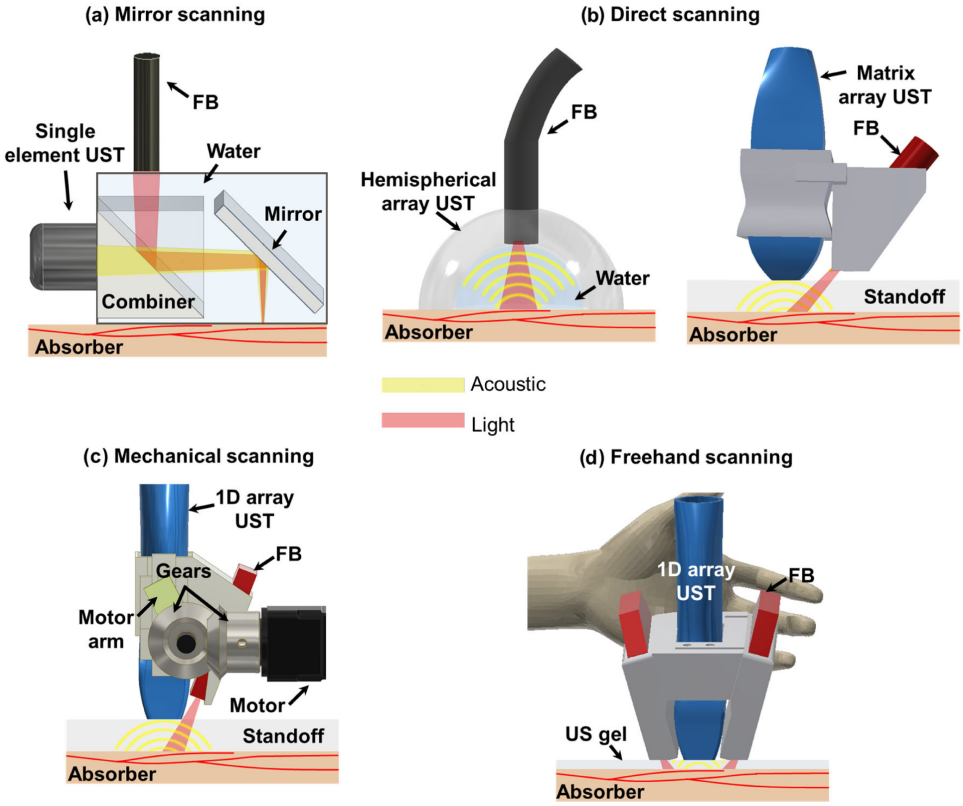


Figure 3.1: Classification of handheld scanning approaches for volumetric photoacoustic imaging as defined by Lee et al. [119]: (a) Mirror scanning employs single-element raster scanning, commonly used in mesoscopy and microscopy. (b) Direct scanning captures a 3D volume using area sensors, matrix arrays, or spherical ultrasound transducers (UST). (c) Mechanical scanning moves a 1D sensor array to sequentially acquire 2D images for 3D reconstruction. (d) Freehand scanning generates 3D volumes through additional processing, with or without external sensors. FB, fiber bundle; US, ultrasound. Figure adapted from [119] under CC BY 4.0 [120].

This or similar classifications are widely used and serve as the structural basis for the following paragraphs, particularly because it has already been established in the closely related field of 3D ultrasound imaging [121]. Notably, this framework also accommodates non-handheld systems, such as robotic-assisted scanning, which can be

classified as a form of mechanical scanning, and stationary scanning setups like the PAMMOTH project for 3D mammoscopy [122], which align with the direct scanning category.

3.1.1.1 Mirror scanning Mirror scanning approaches use a single-element ultrasound transducer in combination with raster scanning techniques to achieve high-resolution photoacoustic images. These methods are typically operated at a different spatial scale, achieving resolutions in the range of several micrometers to a few millimeters. PAI within this resolution range is referred to as photoacoustic mesoscopy, bridging the gap between microscopy and tomography applications [123].

3.1.1.2 Direct scanning Direct scanning approaches enable the real-time generation of 3D photoacoustic volumes using specialized hardware. These systems typically employ integrated array designs to acquire volumetric data without additional probe movement.

One prominent non-handheld example is the Twente Photoacoustic Mammoscope 2 [122, 124], which has incorporated twelve arc-shaped ultrasound arrays into a cup integrated within a bed system. Patients lie on the bed and place the breast into the imaging cup, which can be rotated to acquire multiple projections. This system has been specifically tailored for mammoscopic applications. Another non-handheld device developed by Wray et al. [125] has been designed to capture angiographic volumes of the human extremities. However, this system has been limited to highly localized scanning regions, restricted to specific body parts.

Handheld devices within this category have generally utilized 2D ultrasound array transducers, either in a matrix configuration [126] or with a spherical geometry [127]. Handheld 3D photoacoustic systems have shown promise, particularly due to their potential for wide angular coverage, which can reduce limited-view artifacts [41]. Nevertheless, an evaluation by Neuschmelting et al. [41] has suggested that, despite these theoretical advantages, 2D probes may remain more suitable "particularly for highly absorbing and scattering tissue targets at greater depths with regard to accuracy, reliability and background signaling for spatial orientation" [41].

By design, direct scanning setups capture static 3D volumes, which inherently limit their FOV. Several strategies have been proposed to address this constraint. The 3D PHOVIS toolbox, for example, enables the stitching of multiple 3D volumes into a larger composite image, although manual pre-alignment has been required prior to stitching [128]. In practice, intensity mismatches in overlapping regions have been interpolated,

and overall volumes have been padded or cropped to ensure continuity. The authors have specifically noted that manual alignment was preferred over automated feature extraction due to challenges in processing the inherently sparse and complex nature of 3D PAI data.

Knauer et al. [129] have extended this concept by developing a spatial compounding algorithm based on signal frequency characteristics of vessel patterns in the Fourier domain, rather than relying on spatial features. This innovation has allowed for the generation of large-scale 3D images along arbitrary scanning trajectories, substantially expanding the potential imaging field without sacrificing structural coherence.

3.1.1.3 Mechanical scanning Mechanical scanning approaches constrain the movement of the PA probe, enabling controlled, motorized acquisition over the target area and generally have been independent of transducer configuration. Tomographic slices in mechanical scanning setups have typically been acquired using one of three configurations, as outlined by Fenster and Downey [117] and Mercier et al. [121]: (1) fan-like motion, (2) parallel linear acquisition, or (3) rotation around a central axis.

To ensure consistent scanning speed, these systems have usually incorporated some form of motorization. For example, Kim et al. have leveraged a Scotch yoke mechanism in their system to convert the rotational motion of a motor into linear probe movement [118]. In this setup, the step size between frames has followed a simple harmonic motion, enabling straightforward 3D volume compounding with a scan volume of approximately 40 mm^3 acquired over 20 seconds using three wavelengths.

Similarly, Mozaffarzadeh et al. [130] have demonstrated a motorized scanning system employing a photoacoustic light-emitting diode (LED) system combined with a stepper motor, achieving a step size of 0.054 mm and a scan length between 25 mm and 40 mm with a FOV of 10 mm × 14 mm.

Beyond handheld systems, robotic approaches have been developed to provide controlled probe movement with the added benefit of intrinsic pose tracking, which inherently enables 3D volume compounding. Notably, robotic systems have shown promise in standardizing applied pressure through force control, which is particularly relevant for clinical applications in both US [131] and PAI [132].

However, mechanical systems have faced significant drawbacks, most notably their bulkiness, as exemplified by the 950 g weight of the system developed by Kim et al. [118]. Additionally, because mechanical scanning provides absolute tracking of the probe but not the patient, these systems are unable to capture patient motion, potentially leading to motion artifacts during acquisition.

3.1.1.4 Freehand scanning Freehand 3D US or PAI describes the process in which a clinician manually holds the probe and performs the measurement directly on the patient, while the acquired 2D images are subsequently compounded into a 3D volumetric representation. This process requires that the trajectory of the probe is captured, either using external tracking hardware or by leveraging intrinsic image features, within either a relative or an absolute coordinate system.

Probe tracking Starting with external tracking solutions, optical tracking systems have been widely employed. These systems use stereo cameras to track fiducial markers attached to the probe, which requires a clear line of sight between the camera and the markers. The spatial relationship between the tracked probe fiducials and stationary reference fiducials enables full 6-degrees of freedom (DoF) pose estimation. Some implementations incorporate additional fiducial markers close to the target region to partially compensate for patient motion, addressing a primary limitation of absolute coordinate systems.

This optical tracking approach has been used in PAI by Kirchner et al. [133] for 2D to 3D registration, employing the *Public software Library for UltraSound imaging research* (PLUS) for system calibration [134].

In parallel, electromagnetic tracking systems have been developed to overcome the line-of-sight limitations of optical tracking. These systems generate a low-frequency electromagnetic field around the measurement area, allowing tracking without visual contact. A small sensor placed on the probe detects the local electromagnetic field, from which the pose of the sensor and, consequently, the probe is calculated in real time. Jiang et al. have successfully implemented this approach [135], demonstrating its feasibility for freehand PAI.

Feature tracking In contrast to tracking the probe directly, Sun et al. [136] have proposed tracking skin surface features using a camera attached to the ultrasound probe. As the probe moves, the skin features shift accordingly, and from this relative movement, the probe's pose is estimated. However, as is common in dead-reckoning approaches, a cumulative drift effect has been observed, with positional uncertainty arising from the inherent ambiguity between rotation and translation in video frames.

To address this limitation, two alternative strategies have been developed that circumvent the drift problem. Rafii-Tari et al. [137] have introduced the use of artificial marker strips (checkerboards) for absolute tracking of the ultrasound probe, leveraging a set of collinear points captured by the attached camera to ensure accurate spatial

localization. More recently, Sankepalle et al. [138] have combined an inertial measurement unit (IMU) with visual camera tracking to improve pose estimation stability. They have used the commercial Intel RealSense T265 camera, which integrates a gyroscope, accelerometer, and two fisheye world cameras, enabling 6-DoF tracking by fusing multiple sensor inputs for robust motion estimation.

At the Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference (2024), Dai et al. [139] have presented a novel method conceptually similar to the Tattoo Tomography approach [140], which forms the basis of Part II (Sec. 5). In their work, Dai et al. have developed a three-layered coupling pad containing N-shaped wires in each layer. This design enables pose encoding directly into each 2D ultrasound image, providing 6-DoF positional information based on the spatial arrangement of the wires within the pad.

Reconstruction improvement Beyond advances in tracking precision, recent work has sought to improve freehand and 3D ultrasound reconstruction accuracy by addressing directionality-dependent artifacts. In this context, Göbel et al. [141] have introduced the concept of *Computational Sonography*, which specifically aims to compensate for the direction-dependent intensity variations in ultrasound images. These variations have resulted from different acoustic attenuation paths, causing objects to appear at varying intensities depending on the imaging angle. *Computational Sonography* has offered a framework to account for this effect during the compounding process, ultimately enhancing reconstruction fidelity in freehand ultrasound.

3.1.1.5 Sensorless tracking Speckle decorrelation is one of the fundamental concepts in sensorless 3D ultrasound imaging that has been introduced by Chen et al. in 1997 [142]. This method has utilized the correlation of speckle patterns between consecutive frames to estimate out-of-plane probe motion. Chen has described the origin of speckle as resulting from "the phase-sensitive detection of the scattering from random inhomogeneities in the acoustic properties of biological tissue within the resolution cell of the transducer" [142]. In-plane motion has typically been corrected using optical flow algorithms, such as the method proposed by Farnebäck [143].

In recent years, emerging machine learning-based algorithms have begun to play an increasingly important role in image registration and motion tracking, both in US and PAI. In PAI, deep learning approaches have been used to optimize volumetric image alignment and correct for motion artifacts [144]. In ultrasound, sensorless volumetric imaging has been significantly advanced by extending the concept of speckle correlation

to convolutional neural network (CNN)-based motion tracking [145, 146]. Additionally, more sophisticated methods have leveraged combinations of networks, loss functions, and prior knowledge to improve motion estimation and registration, as demonstrated by Luo et al. [147]. Similarly, Kaderdina et al. [148] have employed a CNN to predict angular displacements between consecutive frames in a constrained setup involving fan-shaped probe movements for infant hip ultrasound. This targeted scenario has led to improved tracking performance and reduced final drift errors compared to more generalized sensorless frameworks.

Unfortunately, these speckle-based tracking approaches are not directly applicable to PAI, since PAI is inherently speckle-free due to its optical absorption-based contrast mechanism [149]. This fundamental difference has presented a major limitation for transferring sensorless techniques developed for ultrasound to the photoacoustic domain.

3.1.2 Multimodal photoacoustic registration

In clinical practice, decision-making is rarely based on a single parameter or imaging modality. Clinicians and clinical guidelines routinely balance the strengths and limitations of various devices, often prioritizing not only diagnostic accuracy but also practical considerations such as device availability, workflow integration, and triage efficiency. Modern interventions, in particular, have increasingly required the fusion of preoperative and intraoperative imaging to guide procedures in real time [52].

Within this context, PAI has emerged as a promising modality that can fill the gap for cost-effective, rapid, and functionally informative imaging. By complementing established gold-standard imaging methods such as MRI and CT, PAI has the potential to provide additional functional contrast without significantly increasing cost or acquisition time (Fig. 3.2).

This section focuses specifically on the role of multimodality in PAI, reviewing advances in image registration strategies and, in particular, recent developments in multimodal PAI registration that aim to spatially align PAI data with other clinical imaging modalities.

A recent review by Chen et al. [9] has showcased significant advances in the field of multimodal PAI and has highlighted the potential of combining PAI with other imaging modalities, such as fluorescence imaging and MRI. These combinations have leveraged shared signal generation and acquisition units or, in the case of MRI, co-registration strategies to spatially align datasets acquired from different modalities.

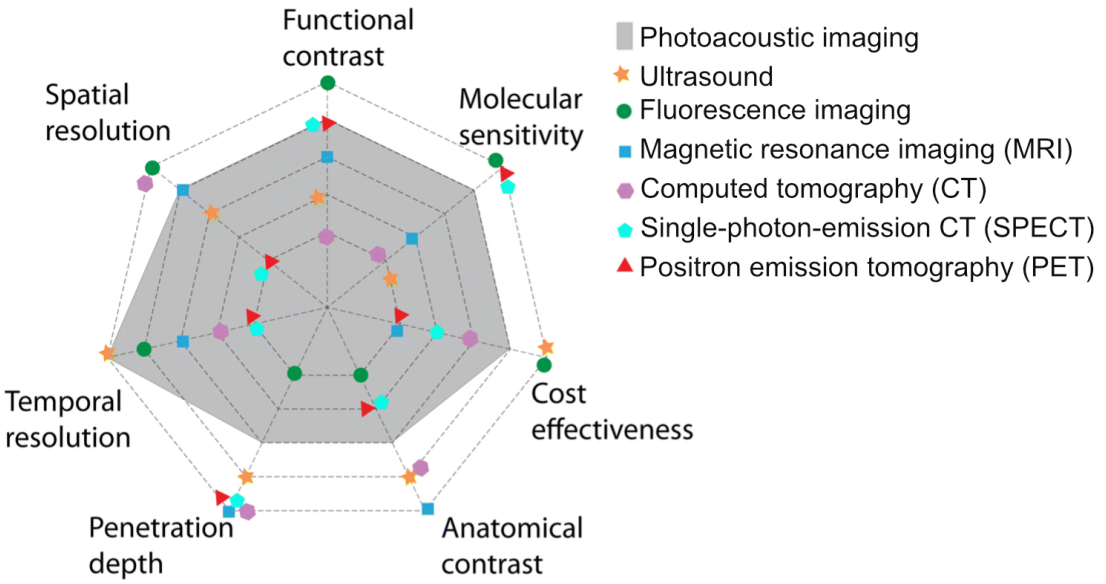


Figure 3.2: Characteristics of different imaging modalities as presented by Chen et al. [9]. The figure visualizes the strengths and complementary advantages of various imaging techniques and emphasizes the need for multi-modality integration in clinical practice. The figure has been cropped and modified (legend adjusted) from the original source [9] and is reproduced under the Creative Commons Attribution (CC BY 3.0) license [150].

The most straightforward and widely adopted multimodal PAI configuration has been the integration of PAI with US using a shared transducer. This co-registration provides the additional benefit of deriving acoustic property maps, such as speed of sound and acoustic attenuation, while also supplying anatomical context that complements the functional information obtained from PAI [151].

Shared optical pathways have also been implemented to combine photoacoustic microscopy with fluorescence imaging [152, 153] or with optical coherence tomography [154]. These hardware integrations have enabled simultaneous acquisition with minimal spatial misalignment, enhancing the complementary value of structural and functional contrast.

In contrast, combining PAI with MRI has required dedicated co-registration solutions, as demonstrated by Gehrung et al. [155]. In their work, a custom mouse bed has been developed to reduce tissue deformation and to mimic the PAI acquisition environment. Their registration pipeline has relied on a two-step landmark-based process. Initially, a rough alignment has been performed using characteristic anatomical features, such

as the spine, followed by a refined landmark registration focusing on the tumor boundaries. This approach has allowed the integration of MRI-based segmentations into PAI processing, including applications like fluence estimation for photoacoustic signal restoration [156].

More recently, Hu et al. [157] have proposed an automatic MRI–PAI fusion algorithm for small animal imaging, using U-Net architectures for MRI segmentation and a transformation network to achieve image co-registration. Kingma et al. [158] have employed a standoff pad containing steel ball fiducials to facilitate the registration of preoperative CT with intraoperative US for human application. In phantom studies, this fiducial-based registration has achieved a mean TRE of 1.84 mm, demonstrating promising accuracy. However, in vivo kidney imaging has shown substantially larger TREs (14.04 mm), primarily due to respiratory motion, indicating that additional refinements such as feature-based registration are necessary for clinical applications.

The challenge of feature-based registration has driven the development of more robust, modality-independent algorithms, such as the modality independent neighbourhood descriptor introduced by Heinrich et al. [159], which has become a widely used method for deformable registration across imaging modalities. This descriptor, later classified by Fu et al. [160] as a deep similarity-based registration algorithm, has provided a device-independent framework for accurate multi-modal alignment.

Within the PAI community, Park et al. [161] have proposed an optical tracking approach (discussed previously in the freehand scanning section) to generate 3D PAI volumes and to perform real-time image fusion of PA/US with pre-recorded MRI images using fiducial markers. This landmark-based registration strategy has achieved a TRE of approximately 4 mm in phantom experiments, providing one of the first demonstrations of real-time PA–MRI fusion in a controlled setting. Both feature-based refinement (Part II Sec. 5.1.2.4) and fiducial-based fusion (Part II Sec. 6) are integral components of this thesis.

3.1.3 Tissue twins in photoacoustic imaging

Multimodal imaging provides the opportunity to combine complementary imaging characteristics beyond the capabilities of a single modality. A logical progression from this is to disentangle imaging limitations from intrinsic tissue characteristics by developing digital models of the tissue of interest, utilizing all available multimodal information.

The concept of super phantoms has been introduced and defined as "phantoms

being able to replicate complex anatomic and functional imaging properties" [162]. These phantoms have often been realized as digital models but are not limited to computational representations. The creation of tissue and device twins, digital models that replicate both biological structures and device specific imaging characteristics, has gained significant attention in PAI. This interest is largely driven by the lack of direct reference for spatially resolved functional parameters, such as tissue oxygenation [50, 163], which are essential for training learning based regression models [40].

In this context, Tang et al. [164] have investigated fluence correction in a digital mouse brain using Monte Carlo simulations. Their work has demonstrated that the strong optical absorption and scattering properties of brain vasculature have introduced substantial disturbances in local fluence distribution, underscoring the need for fluence correction in quantitative PAI.

Similarly, Zhang et al. [165] have constructed numerical phantoms of the human brain using T1-weighted MRI volumes combined with magnetic resonance angiography data. These digital models have enabled the calculation of initial pressure distributions via Monte Carlo simulation and have served as training datasets for U-Net models aimed at correcting skull-induced aberrations in transcranial PAI.

As emphasized, digital twins not only capture biological detail but also enable controlled assessment of device characterization and algorithmic development. In line with this, a recent study by Gröhl et al. [51] has utilized a digital twin framework to systematically compare state-of-the-art photoacoustic reconstruction algorithms.

3.2 Advances in clinical photoacoustics

This section provides a brief summary of relevant clinical studies in the field of PAI, specifically focusing on its application in head and neck cancer treatment monitoring and PAD. In addition, this section compiles a selection of recently identified barriers and roadblocks that currently hinder the clinical translation of PAI. For readers interested in broader clinical applications of PAI beyond the scope of this thesis, several recent reviews can be found, including those covering general clinical settings [166, 167, 168] and oncology-specific developments [169, 170].

3.2.1 Head and neck cancer

A key focus in tumor-investigative PAI studies is the characterization of tumor tissue and the surrounding tumor microenvironment [171]. Becker et al. [172] demonstrated using PAI that reactive and metastatic lymph nodes exhibit distinct chromophore

concentration profiles ex vivo. Additionally, Rich et al. [173] showed in a murine model that PAI can successfully assess hemodynamic changes in head and neck tumors during RT. Furthermore, contrast agent-enhanced PAI has been shown to differentiate malignant from benign lymph nodes in humans [174].

3.2.2 Peripheral artery disease

Karlas et al. [175] conducted a pilot study investigating the ability of 2D PAI to differentiate hemoglobin and oxygen concentration profiles between PAD patients and healthy volunteers (HVs). Their findings demonstrated significant differences in muscle oxygenation between the two groups and further showed that post-surgical restoration of blood flow was associated with normalization of the PAI-derived signals.

Günther et al. [176] presented the first large-scale clinical study (N = 197) evaluating the diagnostic performance of two-dimensional PAI for PAD detection. They confirmed the initial results by achieving an AUROC of 0.839, supporting the clinical relevance of PAI-derived biomarkers for PAD stratification.

Continuing this research, Träger et al. [112] hypothesized that single-wavelength optoacoustic imaging combined with ultrasound could enable mapping of clinical muscle degeneration in progressive stages of PAD. Their hybrid approach aimed to capture both functional and structural muscle alterations associated with chronic malperfusion.

Most recently, Caranovic et al. [177] demonstrated that post-exercise PAI measurements following heel-raise maneuvers could robustly distinguish HVs from PAD patients with high sensitivity and specificity. Their study highlighted the clinical potential of exercise-based PAI protocols to assess PAD-related muscle ischemia in a non-invasive and dynamic fashion.

Within the domain of 3D PAI, research has primarily focused on direct vascular imaging. Choi et al. [178] reported 3D foot vasculature measurements in healthy volunteers using a motorized scanner system, demonstrating the feasibility of high-resolution volumetric imaging. Furthermore, Huynh et al. [179] presented a case study visualizing vascular changes in peripheral vascular disease using 3D foot imaging with a Fabry-Pérot sensor. Their work achieved a 22 mm × 22 mm FOV and illustrated the ability to detect disease-associated vascular remodeling.

3.2.3 Limitations and roadblocks for clinical photoacoustics

To the best of my knowledge, to date PAI is not yet used in any routine clinical imaging procedure, despite the rapidly increasing research interest in the field over the last two decades following the first in vivo PAI experiment [180].

This section provides a brief overview of the major limitations currently hindering the transition of PAI from pre-clinical research to actual clinical application. For a more comprehensive discussion, a recent review summarizing the outcomes of a large community exercise conducted by the International Photoacoustic Standardisation Consortium (IPASC) [6] is recommended.

They identified seven key barriers to clinical translation, which primarily include the small scale and limited generalisability of most current clinical studies together with the fundamental need to demonstrate a measurable positive impact on patient care, as well as a substantial lack of standardization across systems and procedures. In addition, regulatory challenges remain significant, with only a very limited number of Food and Drug Administration (FDA) and Conformité Européenne (CE) certified PAI devices currently available for clinical use. Another persistent obstacle is the inherent difficulty of quantifying PAI data. These points are well-aligned with the conclusions of other recent reviews on clinical photoacoustics [166, 181].

Among these barriers, the unresolved quantification challenge is particularly critical and is consistently mentioned in nearly every PAI study, as a limiting factor for clinical applicability [10, 112, 182].

3.3 Conclusion

Despite significant advances, PAI still faces unresolved limitations that currently hinder its routine clinical adoption. The following conclusions outline how this work addresses the found gaps and contributes to the advancement of the field.

3.3.1 Technical

Conventional PAI systems are predominantly limited to 2D imaging missing 3D spatial and temporal context. This constraint results in several challenges for clinical decision-making, including strong operator dependency, incomplete anatomical coverage, and poor capability for longitudinal assessments due to limited reproducibility. The central hypothesis of this work is that transitioning from 2D to volumetric 3D PAI will substantially enhance the robustness, reproducibility, and interpretability of PAI.

In summary, the technical state of the art presents a variety of approaches for generating 3D PA images, ranging from direct scanning and mechanically constrained movements to tracked freehand scanning. All existing methods, must carefully balance factors such as cost, clinical usability, and image quality, which is essential for seamless clinical workflow integration. To date, no solution has fully addressed the combination of lightweight, device-independent design, high image quality (comparable to 2D devices), relatively large FOV, and minimal calibration requirements. This gap directly limits clinical translation and motivates the technical contribution T1.

A similar gap exists in the field of PA multimodal fusion. Most approaches reported so far have been based on small animal imaging, highlighting the need for further investigation in human-scale multimodal fusion to enable clinical adoption. Here a proof-of-concept for multimodal PAI is presented, integrating the structural imaging strengths of MRI with the functional sensitivity of PAI. Multimodal imaging holds significant potential to overcome key limitations in PAI, including the lack of anatomical landmarks, difficulties in deep tissue visualization, and the absence of standardized reference modalities for benchmarking.

At the same time, biological systems are inherently complex, making it challenging to disentangle true biological effects from imaging-specific artifacts or limitations. Addressing this issue, Manohar et al. recently introduced the concept of “super phantoms” in a perspective article published in *Nature Communications Engineering* [162]. Among the seven open challenges they outlined, two are particularly relevant to this work: (1) multi-modality integrated imaging, and (2) the development of digital twins.

While digital twins have been used to enhance image quality, generate labeled reference datasets, and assess reconstruction algorithms, to the best of my knowledge, simulated PAI images have not yet been applied for failure analysis in clinical PA studies. This work expands the potential of digital twin models by establishing digital twin-based failure analysis, thereby contributing to the evolving concept of super phantoms (T3).

3.3.2 Clinical

There is an urgent need within the PAI research community to "identify important clinical problems where PAI has the potential to either outperform the current standard-of-care, or be more cost-effective” [6]. Addressing this need, two specific clinical use cases are investigated.

The first use case (C1) focuses on the longitudinal monitoring of RT response in

patients. Although promising findings have been reported in the assessment of malignancies in lymph nodes in both murine and human studies, the potential of PAI for longitudinal clinical monitoring of RT response has not yet been explored. This represents a substantial clinical opportunity, especially considering that "there is no established routine of clinical evaluation during RT aimed at assessing early treatment success, particularly with regard to molecular parameters [100, 183]" [93].

Additionally, heterogeneity and malperfusion in the affected calf muscles of PAD patients (C2), is investigated for the first time using 3D PAI. This study aims to demonstrate the clinical usability of the optical pattern (T1) and to directly compare its diagnostic value to that of conventional 2D PAI.

4 Disclosures

Scientific research today is rarely a solo endeavor. It is inherently collaborative and increasingly supported by both human and artificial intelligence tools that enhance efficiency and quality. It is essential to explicitly disclose all contributions and sources of support, both human and computational, that have shaped this thesis.

Beyond the acknowledgments provided in the individual chapters, ideas, methods, and results have been discussed extensively within the weekly Intelligent Systems in Photoacoustic Imaging (ISPAI) group meetings at the German Cancer Research Center (DKFZ) and during my three thesis advisory committee meetings with Lena Maier-Hein, Jürgen Hesser, Jörg Peter and Alexander Seitel. Clinical applications (Part III) have been developed in close collaboration with clinical partners at the University clinics of Heidelberg and Erlangen, who also performed the patient measurements, image segmentations, and provided essential clinical interpretations.

This thesis has been drafted by me, with feedback, discussion, and proofreading support from the ISPAI group, including Lena Maier-Hein. The idea behind the optical pattern represents a group effort, as indicated by the associated patent [184]. Contributions from my prior work in my Master's thesis [185] specifically in Part II Section 5.1 have been clearly distinguished from the novel contributions of this thesis.

The multimodal fusion (Part II Sec. 6) has been conceived and analyzed by me. However, CT measurements were performed by Martina Joachim, Alexander Seitel, and Melanie Schellenberg. MRI measurements were conducted in vivo by Regula Gnirs and on phantoms by Vanessa Franke, Andreas Korzowski, with support from Melanie Schellenberg, who also helped with the PAI measurements and phantom construction.

For the digital twin-based failure analysis (Part II Sec. 7), while I have contributed to the Simulation and Image Processing for Photonics and Acoustics (SIMPA) framework, primarily in processing components such as spectral unmixing, the main simulation framework development has been led by Kris K. Dreher and Janek Gröhl. The optical simulations were jointly performed by Melanie Schellenberg and me, and further refined with Jan-Hinrich Nölke, who also contributed to the acoustic simulations and noise modeling alongside Kris K. Dreher and Christoph J. Bender.

For language enhancement (including: grammar, spelling, and structural refinement) I used large language models, specifically Gemini (Google LLC), ChatGPT (OpenAI Inc.), as well as Grammarly (Grammarly Inc.) and DeepL (DeepL SE). 3D renderings were created using the Medical Imaging Interaction Toolkit (MITK), and for figures and plots I used Matplotlib, seaborn, ChatGPT, and Affinity Designer 2.

- Part II: Technical contributions -

Concept overview

This thesis introduces a framework for *context-aware clinical PAI* that adds essential context along the typical PAI acquisition and analysis workflow (Fig. 5.1). First, a new optical pattern-based approach to volumetric imaging allows acquired 2D slices to be placed into their spatial and temporal context, enabling reproducible 3D reconstructions and longitudinal alignment. Second, by extending the same framework with fiducial markers, PAI volumes can be co-registered with established modalities such as MRI and CT, providing anatomical reference and facilitating multimodal comparisons. Third, digital twin modeling is introduced to add biophysical context, helping to disentangle physiological changes from image formation effects and thereby guiding study design and interpretation. Together, these three dimensions form a compact framework that addresses central barriers to clinical translation and outline a pathway toward making PAI reproducible, interpretable, and clinically relevant.

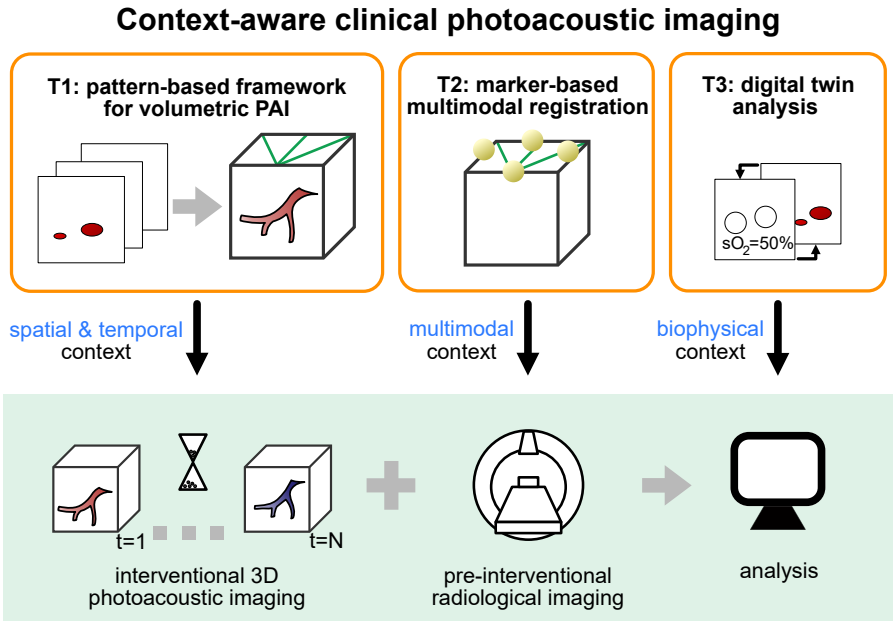


Figure 5.1: Concept overview of context-aware photoacoustic imaging (PAI): The following technical contributions (T1–T3) form a framework that integrates context into the clinical PAI workflow (green) and supports translation into practice. **Spatial and temporal context** (T1) enables reproducible scanning and longitudinal alignment, **multimodal context** (T2) situates 3D PAI alongside established radiological imaging, and **biophysical context** (T3) guides interpretation through a digital twin analysis.

5 Freehand 3D photoacoustic imaging

Analyzing the current state of the art in clinical PAI highlights considerable potential for technical innovation with meaningful clinical impact. Among the most pressing challenges identified is the limited spatial and temporal context inherent to conventional systems (Fig. 1.1).

In response, the following section introduces a novel optical pattern-based approach to 3D PAI. The proposed framework enables reconstruction of a volumetric 3D representation from a sequential stack of 2D photoacoustic slices. Crucially, this slice-to-volume strategy is designed to operate without requiring external tracking systems, complex calibration routines, or substantial hardware overhead. As a result, the approach preserves system portability while expanding the spatial and temporal imaging context, a combination that is particularly relevant for clinical deployment.

Validation of the individual framework components is presented in Section 5.2, where the performance across key technical dimensions is evaluated. This is followed by a critical discussion in Section 5.3, which highlights limitations and outlines future directions aimed at refining the framework's clinical utility and scalability.

5.1 Materials and methods

This section outlines the methodological contribution to the optical pattern-based framework for 3D PAI, partly based on my first author publication [140]. Text blocks taken from this work or my master's thesis [185] is cited and referenced accordingly. The section begins with the general concept and hardware design, followed by a detailed account of the processing pipeline, the practical implementation of the framework, and image acquisition. It concludes with an explanation of the temporal volume registration approach, enabling longitudinal study designs.

While the concept of optical pattern-based 3D imaging was initially introduced in my master's thesis [185], the focus of this section is on the novel developments required to transition from proof-of-concept to a robust framework for clinical translation. A comprehensive validation of the individual system components is subsequently presented in the following section and is further extended in Part III, where the framework is applied to concrete clinical use cases.

Throughout this chapter, in vivo datasets are used to validate, characterize, and exemplify the proposed framework. The key characteristics relevant to understanding the technical contributions are introduced alongside each methodological section. A detailed description of the full datasets is provided in Part III, which is dedicated to

clinical applications. General considerations related to photoacoustic image acquisition are summarized in Section 5.1.3.2 of this part, while the theoretical background of PA image formation is discussed in Part I, Section 2.1.

5.1.1 Optical pattern

Optical pattern-based 3D PAI, also referred to as Tattoo tomography [140], is a patented slice-to-volume reconstruction approach [52], granted in 2025 [184]. It enables spatial pose estimation of the PA probe by leveraging artificial features embedded into the image by means of an optical pattern. By tracking the relative location of individual PA slices, accurate 3D reconstruction of the full acquisition sequence becomes possible.

The complete pipeline is depicted in Figure 5.2, and each pipeline component is elaborated upon in the subsequent sections, with descriptions adapted and extended from the initial implementation presented in [140]:

- **Hardware:** "Prior to image acquisition, an optical pattern is placed above the anatomical region of interest. This pattern should fulfill the following key requirements: (1) It should be designed in a way that a tomographic image of it enables the estimation of the probe pose relative to the coordinate system of the pattern and that (2) it is easy to mount and remove in a clinical setting. (3) The dye that makes up the pattern should absorb in the frequencies that match those of the imaging system (here: near-infrared) and that are complementary to those used for the actual imaging." [140]
- **Image acquisition:** A sequence of 2D PA slices is acquired, each capturing both a portion of the optical pattern and the underlying tissue structures within the region of interest.
- **Pose estimation:** Since the pose is intrinsically encoded by the optical pattern, the spatial configuration of each acquired slice within the pattern reference frame (PRF) can be inferred directly from image features corresponding to the pattern.
- **3D volume compounding:** The final 3D volume is reconstructed by spatially transforming and interpolating the individual slices using the previously estimated poses within the PRF.
- **Temporal 3D volume registration:** In scenarios involving longitudinal imaging where multiple 3D volumes are acquired over time while the optical pattern

remains affixed to the target region a retrospective temporal registration is performed. This process consists of two stages: an initialization step based on the PRF, followed by feature-based registration using robust anatomical landmarks such as blood vessels to account for minor shifts in pattern positioning.

- **Multimodal fusion:** For integration with other imaging modalities (Sec. 6), attachable fiducial markers with known, fixed positions on the optical pattern allow for landmark-based co-registration with external datasets such as CT or MRI.

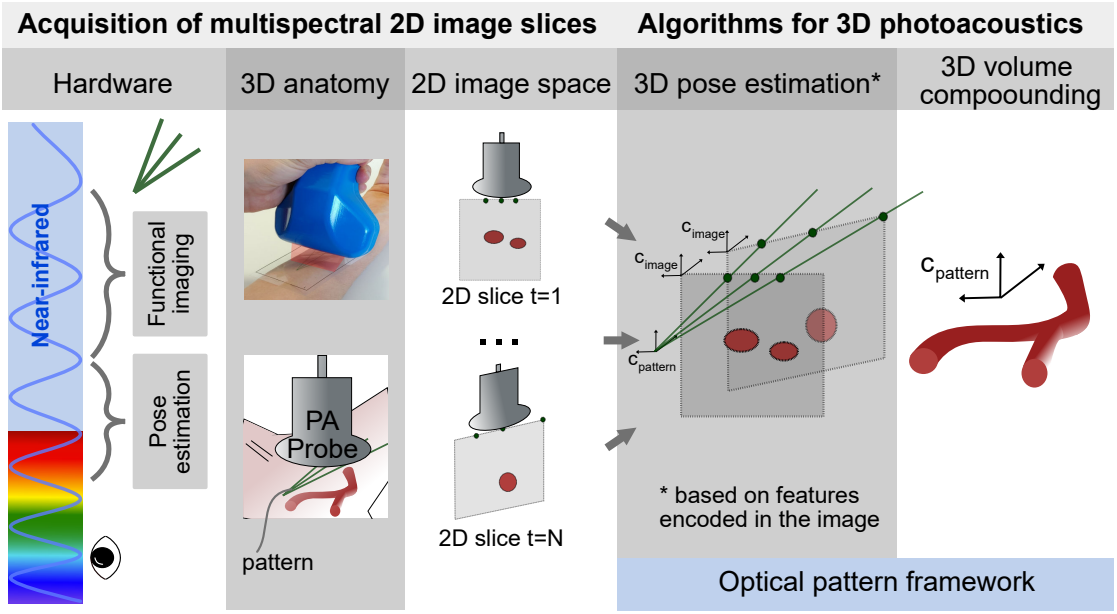


Figure 5.2: High-level schematic of the optical pattern-based pipeline for reconstructing 3D volumes from a sequence of photoacoustic (PA) image slices. The process integrates the physical and algorithmic components required: (**1**) optical hardware setup and pattern design, (**2**) target organ, (**3**) slice-by-slice image acquisition, as well as (**4**) pose estimation via pattern features and (**5**) spatial compounding into a coherent 3D volume as part of the optical pattern framework. Figure adapted from [140] under CC BY 4.0 [120].

5.1.1.1 Hardware Clinical translation of the optical pattern-based approach for 3D photoacoustics is a central aim of this thesis and is reflected in contributions T1 and C2. Applying the technology readiness level (TRL) assessment framework of the Horizon Europe Portal [186], the goal of this work is to advance the hardware components from

TRL ≤ 3 to TRL ≥ 5. This transition marks the step from an initial proof-of-concept to a stage suitable for preclinical and clinical testing, including safety and toxicity assessments.

To support this transition, the following section defines the relevant technical and clinical requirements, followed by a description of the iterative material selection and prototyping process.

Technical requirements The key technical requirements relate to the ink and substrate used in the optical pattern sorted by relevance:

- The ink must exhibit strong absorption below and no absorption above 700 nm, compatible with the 660–1100 nm range of the currently used PAI system to ensure pattern visibility within the PA images for pose estimation.
- The substrate should ideally be optically transparent across the near-infrared window (700–1300 nm) [14], to minimize spectral distortion and allow accurate estimation of chromophores and sO_2 through spectral unmixing.
- The materials should exhibit low acoustic damping to preserve imaging depth and resolution.
- Mechanical compatibility with standard printing techniques and reliable adhesion to the substrate are required.
- The manufacturing process should be cost-effective and scalable to support repeated validation across larger cohorts.

Clinical requirements

- Since the optical pattern comes into contact with patients, biosafety and biocompatibility are essential.
- Consistent visibility of the pattern in 2D PA images directly constrains the usable scan length and thus the achievable FOV.
- Scan quality, particularly in terms of resolution and FOV, strongly depends on scan consistency. Fast scanning speeds or large angular variations can reduce out-of-plane resolution and introduce artifacts.
- For eventual integration into clinical workflows, system usability is critical. The substrate must be easy to apply, remove, and measure.

	Optical Properties	Printing	Clinical requirements
Prototype [140]	×	✓	–
Methylene blue	–	×	–
Inkjet	✓	(✓)	✓
Screen Printing	✓	✓	✓

Table 5.1: Overview of print iterations and corresponding requirements. The prototype does not meet the ink transparency requirement above 700 nm "×". A functional methylene blue pattern could not be printed successfully. The inkjet iteration was neither scalable nor smudge-proof "(✓)". The validation of the remaining requirements is presented in the following section, while entries marked with "–" could not be verified.

Technical implementation Four iterations of the pattern were analyzed (Tab. 5.1). The initial pattern prototype [140] utilized transparency film and standard printer toner. However, this combination lacked the required semi-transparency, limiting its applicability in clinical research. In collaboration with Innovation Lab GmbH (now FLEXOO GmbH, Heidelberg, Germany), several iterations were developed and tested (Fig. 5.3). The validation of these materials against the previously defined requirements is presented in Section 5.2.

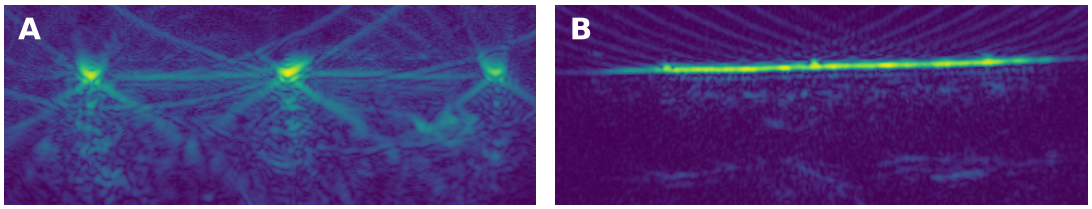


Figure 5.3: Visual comparison of optical pattern visibility using different ink formulations. **A** The original proof-of-concept ink, which is highly absorbing and opaque up to approximately 850 nm, facilitates straightforward peak detection as demonstrated in [185]. **B** The clinically viable ink formulation, while suitable for artifact-free spectral unmixing, necessitates advanced detection strategies, such as background correction, due to the low contrast between the pattern and surrounding skin signal, making peak detection considerably more challenging.

Polyethylene terephthalate (PET*) film (Hostaphan ® RN 100 with 100 μm thickness) was selected as the base substrate due to its widespread availability, mechanical flexibility, and safety profile. Methylene blue was initially evaluated as a dye candidate based on its FDA approval and well-characterized absorption spectrum [187]. However, its

water solubility, poor droplet formation, and low contrast in the printed state precluded further use.

In the next prototype iteration, 0.06 wt% (mass fraction) of Macrolex® Blue 3R FG (Lanxess, Cologne, Germany) [188] was selected as a suitable ink. This dye offers broad applicability in inkjet formulations, with reported ∼90% transmission above 700 nm and increased absorption at shorter wavelengths. The ink was inkjet-printed, resulting in 16 usable patterns that were deployed in a clinical pilot study.

To address issues of scalability and print durability, the subsequent iteration was conducted using screen printing (Fig. 5.4). A thermal dielectric (Dupont Micromax ® PE773) served as the encapsulent with the same Macrolex® Blue 3R FG dye. This improved approach yielded over 300 usable pattern films, enabling extended technical validation and use in a subsequent clinical study (Part III).

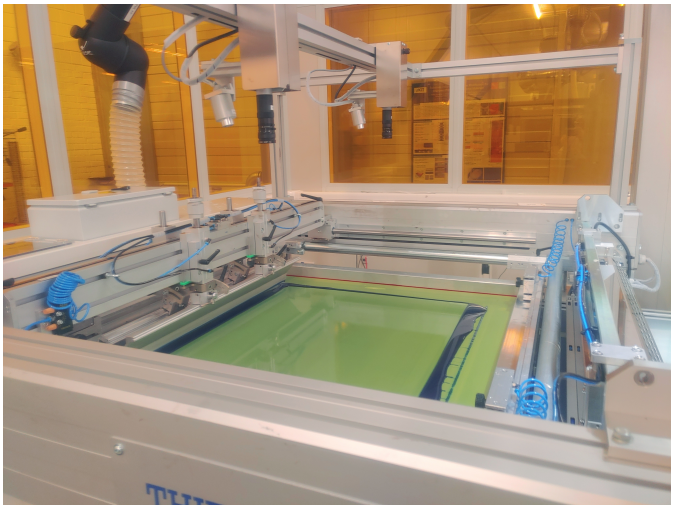


Figure 5.4: Schematic of the screen printing process used to fabricate the optical pattern. The pattern was printed using a THIEME 3010 industrial screen printer (THIEME GmbH & Co. KG, Teningen, Germany). During the printing process, a screen containing the predefined pattern is mounted in the printer. The dielectric-dye mixture is applied on top of the screen, while the substrate is positioned beneath. As the squeegee moves across the screen, the dye is pressed through the screen opening, transferring the design onto the substrate.

Design "For this first prototypical implementation, a trident design was chosen, as depicted in Fig. 5.5. This trident consists of an isosceles triangle (for simplicity referred to as tilted lines) and the corresponding angle bisector between the tilted lines of the

triangle (referred to as central line). Mathematically speaking, this initial prototype pattern design currently only allows accounting for three degrees of freedom (DoF) [...]. The remaining three DoF are recovered by imposing constraints on the image acquisition process" [140].

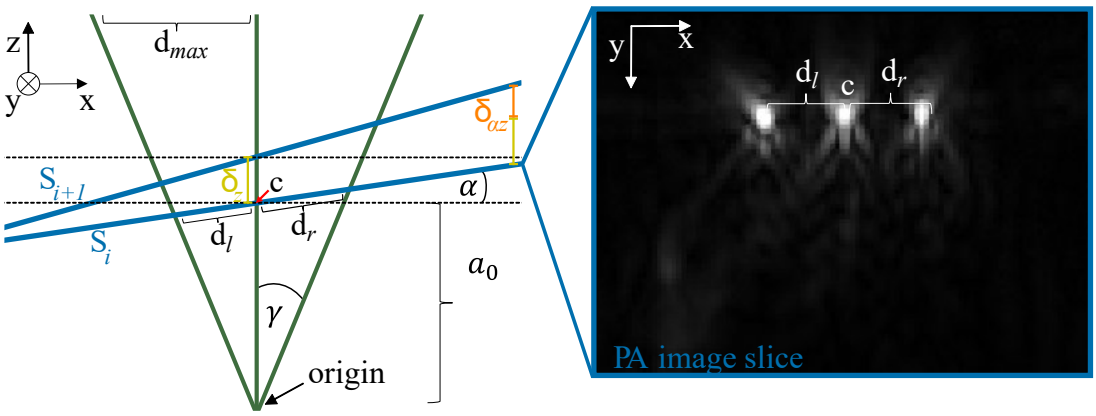


Figure 5.5: Schematic overview of the optical pattern (green) in the x-z plane, intersected by a photoacoustic image slice S (blue). The coordinate system (x: width, y: depth, z: scan direction) is consistently used throughout this thesis. All relevant pattern variables are illustrated, including the distances d_l and d_r, the central intersection point c, the pattern opening angle γ, the pattern width d_{max}, the distance of the slice to the origin a_0, and the rotation angle α. The displacement δ_z between consecutive slices S_i and S_{i+1} is indicated in yellow, while the additional rotational offset $\delta_{\alpha z}$ caused by differences in α between slices S_i and S_{i+1} is shown in orange. Figure adapted from [140] under CC BY 4.0 [120].

5.1.1.2 Constraints The pattern design involves several constraints and important considerations during scanning.

(a) Orthogonality constraint: By design, the pattern enables pose recovery with respect to three DoF: rotation around the y-axis and translations along the x- and z-axes (Fig. 5.5). This means that movements involving rotation around the x-axis, fan-like angular sweeps, are not recoverable and therefore must be avoided. Operators must maintain probe orientation orthogonal to the pattern surface throughout the scan. A more detailed discussion of the impact of orthogonality violations on the reconstruction process can be found in Chapter 4.2.1 ("Scale Analysis") of [185], which concludes: "PAI is most sensitive the closer an image pixel is to the surface. Therefore, a tilt of the

PA probe with an angle ξ [(rotation angle around x axis)] does not have a remarkable influence on the most features of interest" [185].

(b) Flat surface constraint The current implementation of the optical pattern framework assumes a flat imaging surface. This assumption is inherent to both the pattern geometry and the peak detection algorithm, where the shortest detected distance between peaks is used for coordinate computation. While this approach simplifies reconstruction and enables efficient pose estimation, it also imposes geometric constraints on the applicability of the method in anatomically curved regions.

(c) Scan speed and resolution: The lateral resolution in the final 3D volume is influenced by the scan speed v_z, the effective sampling frequency f_s, determined by the total number of wavelengths n_{tot} in the preset (Eq. 5.1), and the slice thickness in the 2D PA image, hence the out-of-plane resolution, as well as the frequency of occurrence of the pattern wavelength 665 nm in the preset $n_{665\ nm}$.

$$f_s = \frac{25\ \text{Hz}}{n_{tot}} \times n_{665\,\text{nm}} \tag{5.1}$$

Additionally, angular displacement δ_z (Fig. 5.5) and its rate of change (angular velocity) affect the local scan speed at the edges of the imaging plane, ultimately limiting the achievable resolution. To avoid undersampling and satisfy the Nyquist criterion (Part I Sec. 2.2.3 [66]), the operator must maintain a scan speed v_z that accounts for both linear and angular displacements.

(d) Pattern visibility: The visibility of the optical pattern within each 2D PA frame is essential, as it forms the basis for deriving pose information. For extended scan lengths or when features of the pattern are widely spaced, some elements may fall outside the imaging FOV. This imposes a hard constraint on the maximal usable scan length and thus defines the effective 3D FOV. Accordingly, pattern geometry must be designed with this limitation in mind.

5.1.2 Algorithmic approach for 3D reconstruction

Following acquisition and pre-processing (Sec. 5.1.3.2), the resulting input of the optical pattern framework consists of a stack of 2D photoacoustic images acquired at multiple wavelengths, including the encoded probe pose at 665 nm, along with a co-registered

stack of ultrasound images. The following sections describe the methodological contributions enabling the reconstruction of a 3D volume from this multi-slice dataset using the device-independent optical pattern-based approach.

5.1.2.1 Detection of the optical pattern The first essential step in slice-to-volume registration using the optical pattern-based method is the detection algorithm. Its primary objective is to robustly identify absorption peaks within each 2D PA image slice, specifically three for the current trident pattern design.

Background correction The newly developed ink designed to meet the requirement of optical transparency above 700 nm, exhibits weak absorption even at the target wavelength of 665 nm. Although the optical properties of the *new* ink represent a substantial advancement toward clinical applicability, enabling artifact-free spectral unmixing, it also poses a significant challenge in terms of pattern detection. The reduced absorption compromises the visibility of the pattern in the 2D PA images, affecting both the image acquisition process (requirement of live visibility of the pattern Section 5.1.3.2) and the automated peak detection procedure within the pose estimation algorithm.

Inspired by techniques from astronomical image processing and digital photography, where dark-frame subtraction is commonly used to reduce noise, a background correction method is introduced to suppress tissue-related signals and enhance the visibility of the optical pattern. This approach is visualized in the validation section (Fig. 5.15).

As described earlier, the absorption properties of the pattern dye are designed such that it becomes optically transparent above 700 nm. However, a major challenge arises from the fact that biological tissue, and particularly skin and melanin in close proximity to the optical pattern, absorbs strongly at 665 nm, the wavelength used for pattern detection. This overlap in absorption spectra makes it difficult to distinguish the pattern signal from that of the surrounding tissue. To address this, a functional PA image at a wavelength above 700 nm (where the pattern is not visible), denoted as $\mathrm{Img}_{\mathrm{func}}$, is subtracted from a pattern-visible image at 665 nm, denoted as $\mathrm{Img}_{\mathrm{pat}}$, resulting in a residual signal S_p that predominantly contains the pattern information:

$$\mathrm{S}_p = \mathrm{Img}_{pat} - \mathrm{Img}_{func} \cdot \delta_S. \tag{5.2}$$

Here, the correction factor δ_S compensates for differences in signal intensity between the two wavelengths, allowing the resulting S_p to be optimized for maximal contrast

and improved peak detectability:

$$\delta_S = \min\left(\sqrt{\left(\mathrm{Img}_{\mathrm{func}} - \mathrm{Img}_{\mathrm{pat}}\right)^2}\right). \tag{5.3}$$

The value of δ_S is determined from cropped image regions centered around the skin ($\pm$30 pixels). Optimization is carried out using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [189] as implemented in SciPy (v1.14.1) [190], in combination with the RMSE function from scikit-learn (v1.6.0) [191].

If the BFGS optimizer fails to converge, an alternative fallback strategy is applied. A custom fit parameter is iteratively adjusted in increments of 0.1 within a range of $\pm$50. The final correction factor is selected as either (1) the largest value for δ_S which the median of S_p is negative, or (2) the default value of 1 if no such parameter is found. This fallback logic is detailed in the pseudocode provided in Algorithm 5.1.

Algorithm 5.1 Estimation of δ_S based on median difference

```
Set δ_S to 1
Compute initial difference: dif ← Img_pat − Img_func
while median of dif > 0 do
    Increment δ_S
    Update dif based on scaled Img_func
    if δ_S exceeds upper bound then
        Set δ_S to 1
        Exit loop
    end if
end while
while median of dif < 0 do
    Decrement δ_S
    Update dif based on scaled Img_func
    if δ_S exceeds lower bound then
        Set δ_S to 1
        Exit loop
    end if
end while
```

In addition to the global intensity variations captured by the scaling factor δ_S, the photoacoustic signal is also subject to nonlinear relative changes due to the distinct absorption spectra of the constituent chromophores. Therefore, the background correction method described in the previous section is most effective when the anatomical features exhibit similar signal profiles at both 665 nm and the selected functional

wavelength w_{func} used in Img_{func}. In practice, presets (lists of selected wavelengths) typically include at least one wavelength above 700 nm and below the isosbestic point of hemoglobin near 800 nm. From these, the wavelength w_{func} is selected to maximize background similarity with the pattern-visible image:

$$w_{\text{func}} = \min \{w - 665\,\text{nm} \mid w \in \text{Preset} \setminus \{665\,\text{nm}\},\ w - 700\,\text{nm} > 0\}\,. \tag{5.4}$$

To further enhance the contrast of the pattern signal S_p, a sharpening kernel is applied to the background-corrected image using OpenCV's `filter2D` function [192]. Subsequently, all pixel values below a defined threshold, set as the image's standard deviation, are zeroed to suppress residual noise.

Detection The resulting enhanced image serves as input to the peak detection algorithm previously described in [185], implemented using the `find_peaks` method from SciPy. This method identifies the n most prominent pixels in the image, where n corresponds to the number of expected pattern peaks. Prominence is determined not only by absolute intensity but also by local contrast relative to neighboring pixels. This strategy ensures that only a single representative pixel is selected per peak. Conceptually, each peak corresponds to the intersection point between the PA image plane and the trajectory of one of the trident pattern lines.

To improve precision, a novel refinement step is applied to each detected peak. A local image patch (by default, 6 pixels wide) is extracted around the candidate location. The intensity profile of this patch is projected onto the x- and y-axes, and a Gaussian curve is fitted to each projection. If the fit converges, the mean of the Gaussian is used to update the peak position.

5.1.2.2 Pose estimation Following refinement, a line registration step is performed as described in [185]. The registration process imposes a constraint that promotes continuous line structures, which helps to reduce the number of falsely detected peaks, such as high-intensity artifacts (e.g., nevi). Its purpose is to evaluate whether the detected peaks belong to the optical pattern, based on the certainty values obtained from the `find_peak` method. This is achieved through an iterative approach that assigns each detected peak to a line array. According to the design of the pattern, $n = 3$ lines are expected to be identified. As an initial estimate, the slice with the highest total certainty is selected. Subsequently, the peaks in adjacent slices are examined. If the x-position of a peak lies within a defined tolerance (*peak_thresh* Tab. 5.2) relative to

the neighboring peaks, it is assigned to the same line; otherwise, a new line array is initiated. Once all peaks have been assigned to line arrays, the n most likely lines are selected based on the number of assigned peaks and the combined certainty associated with each line.

With the final refined peak positions, the spatial coordinates of the optical pattern are determined, enabling the subsequent registration of the PA image slice into the global pattern coordinate system.

The optical pattern with the trident design "is characterized by the distance a_0 of the absorption peak of the central pattern line c to the pattern origin and the slice tilt angle α as defined in Fig. 5.5. These are computed as follows:

$$tan(\alpha) = \frac{1}{tan(\gamma)} \cdot \frac{d_l - d_r}{d_l + d_r}, \tag{5.5}$$

$$a_0 = \frac{(d_l + d_r) \cdot cos(\alpha) + \frac{(d_l - d_r)^2}{d_l + d_r}}{2 \cdot tan(\gamma)}, \tag{5.6}$$

where d_r and d_l are the distances between the right and left point to the central point, respectively, and γ is the pattern opening angle. When assuming an orthogonal pose of the probe relative to the pattern plane, the line given by α and a_0 uniquely determines the pose of the probe in 3D. This results in the following transformation matrix $T_{img \rightarrow tattoo}$, which transforms the PA image slice into the tattoo [optical pattern] coordinate system:

$$T_{img \rightarrow tattoo} = \begin{pmatrix} cos(\alpha) & 0 & sin(\alpha) & -x_c \\ 0 & 1 & 0 & -y_c \\ -sin(\alpha) & 0 & cos(\alpha) & a_0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{5.7}$$

where x_c and y_c are the x and y coordinates of the central *Tattoo* point c in the PA image slice and α and a_0 the previously derived *Tattoo* coordinates" [140].

In practice, the derived set of spatial coordinates across a scan is smoothed using a Savitzky–Golay filter. This is based on the assumption that the probe motion during scanning is continuous and smooth, whereas the initially detected coordinates are derived from discrete pixel locations and therefore subject to quantization artifacts [185].

At this stage, coordinates have only been computed for those frames within the PA scan that contain visible pattern information. However, the acquisition preset typically includes multiple additional functional PAI frames that do not contain detectable

pattern features. To obtain transformation information for these intermediate frames, linear interpolation is applied between the two temporally closest frames with valid coordinate data. This yields an estimated pose for each frame across the entire scan. The complete set of coordinates is then stored in a dedicated dataframe for subsequent processing.

While the transformation (Eq. 5.7) of each 2D slice into the reference frame of the optical pattern is of interest, the ultimate objective is to compound these slices into a coherent 3D volume. This process is the focus of the next chapter.

5.1.2.3 Compounding After the pose of each 2D PA slice has been determined within the global coordinate system (Fig. 5.5), defined by the pattern with the coordinates α and a_0, each pixel of these slices is mapped into a standardized volume. This volume is initially set to 1.5 times the size of the pattern and is later cropped to the pattern dimensions and with the derived skin level at a height of 50 pixels. The spacing in the x–y plane is set according to the input slice spacing, while the z-spacing can be adjusted based on scan speed and out-of-plane resolution (Sec. 5.2.1.2).

The compounding technique employed in the optical pattern-based approach is a pixel-based method, specifically pixel nearest neighbor with hole-filling from a local neighborhood (Part I Sec. 2.2.3). In this strategy, each pixel from the 2D PA slices is mapped to its nearest voxel in the 3D data volume (distribution step), while simultaneously incrementing a corresponding counter volume. If multiple pixels fall into the same voxel, their values are averaged [185].

Due to rapid probe movement or angular changes during scanning, some voxels may remain unfilled after the initial compounding step. These empty voxels represent gaps in the volume data that must be interpolated to ensure spatial continuity. The previous implementation relied on linear interpolation along the scan direction using `scipy.griddata`, but this approach proved to be a significant bottleneck in overall processing time [185]. To address this limitation, a new, computationally efficient interpolation routine was developed.

Hole-filling generally consists of three main steps (1) identifying empty voxels, (2) computing interpolated values, and (3) assigning those values back into the correct voxel locations.

1) Hole identification: Leveraging the earlier voxel distribution step, a *data volume* containing accumulated pixel intensities and a *counter volume* tracking the number of pixels per voxel are divided to produce an averaged volume. Voxels with zero

counts in the counter volume are set to `NaN`, marking them as unfilled [185].

2) Interpolation: To perform linear interpolation along the scan (z-) direction, a missing voxel value in slice i is estimated as the mean of the corresponding voxel values in the adjacent slices $i-1$ and $i+1$. Formally, this can be written as:

$$V_i(x,y) = \frac{1}{2}\left[V_{i-1}(x,y) + V_{i+1}(x,y)\right], \tag{5.8}$$

where $V_i(x,y)$ denotes the voxel value at position (x,y) in slice i. Operationally, this is implemented by adding the original volume with itself shifted forward along the z-axis by two pixels, and dividing the result by two. This produces a new interpolated volume containing potential fill values for slice positions that have valid neighbors on both sides. Regions with two or more consecutive missing voxels (in z-direction) remain unfilled, thereby preserving the structural integrity of sparsely sampled regions.

3) Assignment: The interpolated values are inserted back into the original volume at the locations where `NaN` entries were previously identified. This results in a lightweight linear interpolation method that maintains spatial coherence without attempting to overfill deeply undersampled regions.

While this method does not recover all missing voxels, it achieves a balance between speed and accuracy. As discussed in [59], many pixel-based methods incorporate a predefined maximum hole-filling distance to limit interpolation artifacts. The primary performance gain of the proposed method stems from using efficient matrix operations in both the hole detection and interpolation steps.

Finally, users are provided with flexible options for data output (Tab. 5.2). The compounded data can be stored either as N multispectral 2D frames, as separate 3D volumes for each wavelength, or as a single four-dimensional volume, with spectral information encoded along the fourth dimension.

5.1.2.4 Temporal volume registration In the preceding sections, we have established a method that enables the integration of spatial context into PAI. However, clinical workflows frequently extend beyond single acquisitions, relying instead on longitudinal measurements to monitor tissue evolution over time. Analyzing repeated measurements requires distinguishing between anatomical shifts (e.g., changes in probe position) and physiological or pathological tissue responses. However, in practice, longitudinal

acquisitions, whether due to reapplication of the optical pattern, probe replacement, or repeated measurements, rarely result in identical imaging conditions. Small variations such as changes in probe pressure, minor displacements of the pattern, or patient movement introduce misalignments between repeated volumetric scans.

Of particular concern are differences in applied probe pressure, which influence both the signal amplitude and anatomical positioning in the reconstructed volume. These effects typically lead to non-rigid deformations, making spatial registration across time points especially challenging. Similar challenges have been widely studied in the field of ultrasound imaging. Model-based techniques have been proposed to correct for tissue deformation caused by varying probe forces, including robotic setups that directly measure and model applied force [193, 194].

The following section demonstrates how the existing optical pattern-based framework can be extended with an additional image-based registration step to support temporal context integration. Specifically, a temporal registration algorithm is introduced as a post-processing step. This method automatically aligns two PA volumes using a rigid transformation to mitigate displacement artifacts (Part I Sec. 2.2.2).

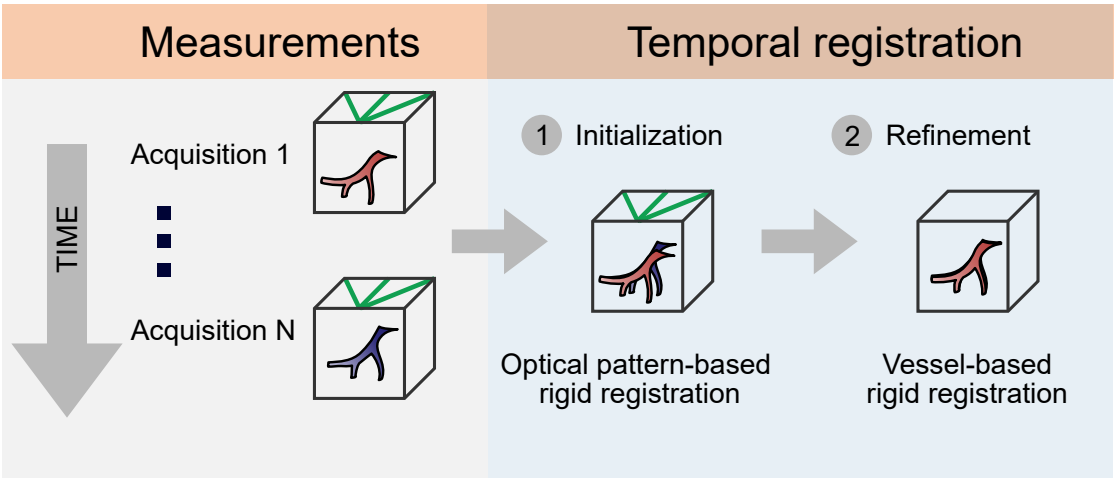


Figure 5.6: Schematic illustration of temporal image registration across multiple longitudinal measurements. Natural anatomical landmarks - such as blood vessels - may vary slightly in location between acquisitions. The registration pipeline begins with an initialization step based on the optical pattern (1), followed by a refinement step using high-intensity vessel structures (2).

The core idea of the proposed temporal registration method is to perform rigid alignment of 3D PA volumes acquired at different time points using prominent high-intensity anatomical landmarks (Fig. 5.6). As an initial alignment, a fiducial registration

based on the optical pattern is applied. This initialization represents the expected outcome in an ideal laboratory setting with perfect pattern reapplication and stable probe placement.

To refine this alignment, anatomical landmarks must be extracted from the volumetric data. Ideally, these landmarks are automatically detectable to ensure robustness and reproducibility. In the context of PAI, blood vessels represent the most prominent high-contrast structures beneath the skin surface, making them ideal candidates for use as landmarks in registration. For landmark detection, a thresholding approach is applied to the 800 nm volumes. Specifically, a binary mask is generated for both the fixed and moving images by selecting voxels exceeding the 98th percentile of intensity values beneath the skin surface. The skin level, as well as the initial transformation, are known from prior processing within the optical pattern-based framework. Finally, a rigid transformation between the fixed and moving volumes is computed using the `SimpleITK` Python toolkit [195], completing the registration process. A real-world application of this registration pipeline, including an exemplary evaluation of its performance, is presented in the upcoming validation section (Sec. 5.2.2.4).

5.1.3 Implementation

This section outlines the practical implementation of the pattern framework and the corresponding image acquisition protocol.

5.1.3.1 Framework The reconstruction framework provides an interface for *processing* (Fig. 5.7), which serves as the primary access point. It is structured into two main modules: (1) pose estimation and (2) compounding. These core modules will be described in detail in the subsequent sections.

In addition to usability and accessibility, computational runtime plays a critical role in the user experience and the translational potential of the system. Currently, due to manufacturer-imposed restrictions, direct live access to raw data via an application programming interface (API) is not available. Nonetheless, achieving short and efficient processing times is important, with the long-term perspective of enabling device integration. Accordingly, the upcoming validation not only demonstrates technical accuracy but also benchmarks processing time against the duration of the corresponding PA data acquisition, offering a qualitative measure of performance.

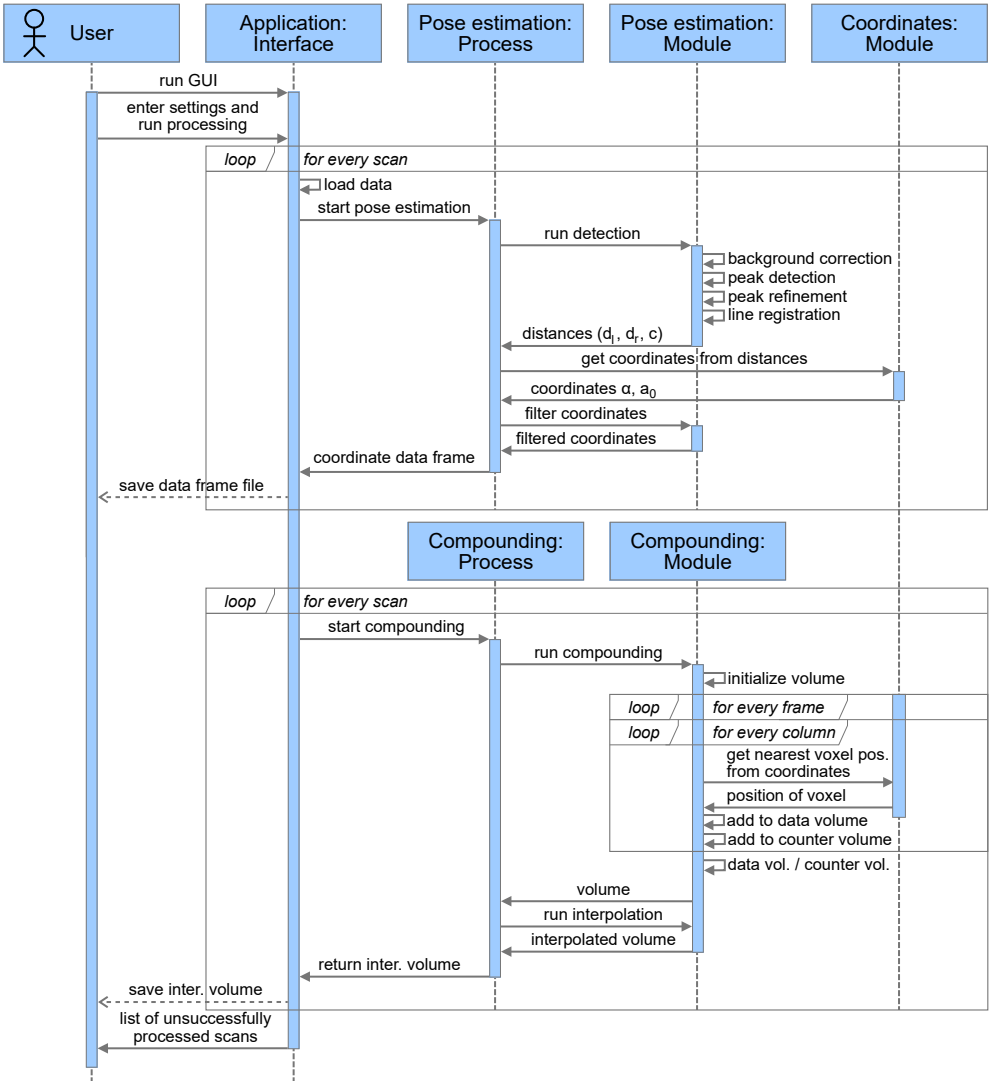


Figure 5.7: Unified modeling language (UML) sequence diagram of the optical pattern framework for 3D photoacoustic imaging (PAI). The process starts with the user interacting through the application interface to initiate pose estimation and compounding, providing the necessary settings (Tab. 5.2). The pose estimation module performs background correction, peak detection and refinement, line registration, and finally returns the (filtered) coordinates (Fig. 5.5). The compounding module then maps the pixel data to the nearest voxel in the 3D volume based on these coordinates, followed by interpolation. Errors raised throughout the pipeline are captured, and the corresponding scan ID is returned to the user upon completion. Logging and verbose plotting are not shown in the sequence diagram for simplicity.

Setting	Value	Comment
save_as_2D	true	Save output as 2D (+wl) frames
save_as_3D	false	Save output as 3D volumes with each wavelength as a separate volume
save_as_4D	true	Save output as 4D volumes with wavelength as fourth dimension
do_inter	true	Perform interpolation
verbose	true	Print detailed logs during processing
expected_lines	3	Expected number of pattern lines
spacing	[0.1, 0.1, 1]	Spatial resolution of input in x, y, and desired resolution z directions
p_length	50	Pattern length
p_width	30	Pattern width
tolerance	1	$\alpha = 0$, if $\lvert \text{dist}_l - \text{dist}_r \rvert \leq$ tolerance [pix]
filter_param	dist	Apply filter to distances or coordinates
filter_type	savgol	Type of filtering: Savitzky-Golay or false
reg_alg	2	Version of registration algorithm
kernel	11	Kernel size used for filtering operations
flip	0	Left/right flipping of input images/data (0: no, 1: yes)
layout	trident	Pattern layout configuration
wavelengths	[665, 760, 850]	List of illumination wavelengths in nm
fit	minimize	Fit method used for background correction: minimize or median
bg_corr	true	Perform background correction
d_min	2.5	Minimal allowed distance between pattern features - cropping in z-direction (from origin)
first_sweep_id	null	Time offset / cropped imaging sequences
show_plots	false	Verbose visual output while processing
skin_level	143	Approximate skin height in pix
pose_crop	[0, 0, 0, 0, 0, 0]	Cropping parameters [x_{min}, x_{max}, y_{min}, y_{max}, z_{min}, z_{max}]
nawl	[665]	Wavelength with detectable pattern
it_bf	true	Use iThera provided reconstruction
relevant_data_json	null	Dict with (subset) of scans for processing
type_array	[["PA", "all"], ["SU", 800, "Hb"]]	Specify which data used for compounding
peak_thresh	10	Max. distance in x-direction [pix], that two consecutive peaks are considered part of the same pattern line

Table 5.2: Settings for the optical pattern-based 3D reconstruction framework. Emerald rows are used for pose estimation, apricot rows for compounding, and white rows are relevant to both.

For ease of use and streamlined access to the processing pipeline, a functional *Tkinter* graphical user interface (GUI) [196] was developed (Fig. 5.8). The GUI enables users to load an arbitrary number of 2D PA scans, configure relevant processing settings (Tab. 5.2), and directly invoke the pose estimation and compounding modules to generate volumetric reconstructions. A system-level test demonstrating the functionality of the complete processing pipeline is provided in the subsequent validation section.

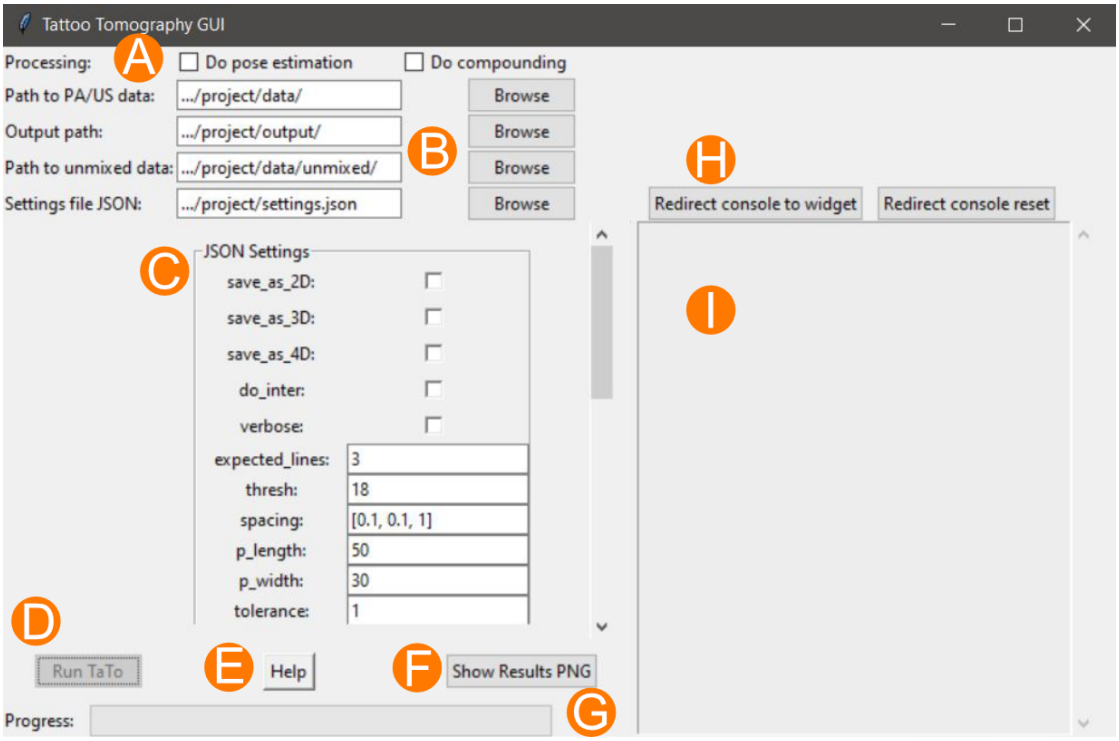


Figure 5.8: Graphical user interface (GUI) of the optical pattern-based framework. (A) Selection of processing modules, (B) input/output path configuration, (C) adjustment of processing settings (Tab. 5.2), (D) execution of the processing pipeline, (E) user guide, (F) visual inspection of results, (G) progress bar, (H) redirection of console output to the GUI for monitoring processing status, and (I) embedded console canvas for runtime messages and logs.

5.1.3.2 Principles of image acquisition with an optical pattern A fundamental understanding of the image acquisition and processing workflow is essential to fully appreciate both the strengths and limitations of the optical pattern-based approach for 3D PAI. This section focuses on the specifics of image acquisition using an optical pattern, as well as the general PA image processing pipeline. While the optical pattern-

based framework for 3D PAI is, by design, device-independent, all datasets used in this thesis were acquired with a MSOT Acuity Echo system (Part I Sec. 2.1.3).

The PAD study (Part III Sec. 9) employed the CE-marked version of the system, while all other experiments were conducted using a non-CE version for research purposes. Although both system variants are nominally identical in terms of technical specifications, the CE version includes an extended self-test protocol. It is worth noting that device-specific variability can introduce systematic bias [197]. Differences between individual devices arise from manufacturing tolerances, maintenance history, and variations in transducer coatings. For instance, each probe version is equipped with a distinct coating material, and even minor shifts in membrane positioning have been observed following routine service interventions. Such variability underscores the importance of device-aware interpretation when aggregating or comparing datasets.

The following acquisition workflow (Fig. 5.9) is described in the context of the MSOT Acuity Echo, as introduced earlier. While the optical pattern-based approach is in principle device-agnostic, other handheld PAI devices from iThera Medical, third-party vendors, or custom-built systems may require additional adaptations to the acquisition protocol.

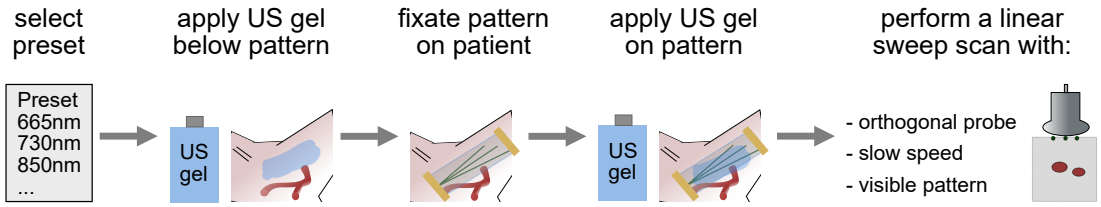


Figure 5.9: Acquisition pipeline for the optical pattern-based scanning. The process includes preset selection, ultrasound (US) gel application below and on the pattern, pattern fixation, and a linear sweep scan performed with orthogonal probe positioning, slow scan speed, and ensured pattern visibility.

The acquisition process begins with the selection of a wavelength sequence, referred to as a *preset*. On the research-grade MSOT system, this configuration can be adjusted by the operator. In contrast, the CE-marked system mandates certification for each preset, which must therefore be generated by authorized iThera Medical staff. Presets typically include wavelengths around the isosbestic point of hemoglobin (approximately 800 nm) to enable blood oxygenation estimation. As outlined previously, the dye used in the optical pattern exhibits increasing absorption toward shorter wavelengths. To ensure that the dye is visible in the PA images, at least one wavelength shorter than 700 nm must be included in the preset. Empirically, it was observed that the MSOT

system exhibits unstable laser pulse energy at its shortest tunable wavelength (660 nm), frequently triggering shutdowns, especially in the CE version, where an internal safety switch deactivates the system if laser energy fluctuates beyond a defined threshold between consecutive frames. As a result, a wavelength of 665 nm was selected as a stable alternative for achieving high pattern visibility.

For measurements using the optical pattern, the printed film must be affixed to the patient using e.g., medical-grade plaster. To ensure acoustic coupling, ultrasound gel is applied between the skin and the film before fixation, and a second layer is added on top of the film immediately prior to scanning. The fidelity of the reconstructed 3D volume depends critically on both the quality of the acquired 2D PA images and the accuracy of the derived probe position [59]. During acquisition, the operator must perform a controlled linear sweep of the probe across the surface of the optical pattern while adhering to three primary constraints: orthogonality, scan speed, and pattern visibility in the 2D image (Sec. 5.1.1.2).

5.1.3.3 Image processing Following acquisition, several image processing steps are required to convert the raw PA time series data into 2D reconstructed images. The MSOT system provides three data streams: a 2D US image, the PA time series signals, and an estimate of the laser pulse energy for each frame.

To correct for energy fluctuations between repeated measurements at the same wavelength and across different wavelengths, the PA time series data are normalized by the corresponding laser energy. The energy profile of the MSOT laser source is wavelength-dependent: it is lowest at shorter wavelengths, peaks around 730 nm, and then decreases monotonically toward longer wavelengths.

Systematic hardware effects, such as sensor offsets, sensitivity variations, and bandwidth limitations, are reduced through bandpass filtering. A typical filter is configured with a passband from 0.1 kHz to 8 MHz and shaped using a Tukey window with $\alpha = 0.5$.

The transformation from the time domain to the spatial domain is carried out using DAS beamforming [23]. After beamforming, envelope detection is applied using the Hilbert transform to eliminate negative signal components. As defined in [198], the Hilbert transform of a signal $g(t)$ is the convolution of $g(t)$ with $1/\pi t$.

Finally, linear spectral unmixing is used to estimate the relative concentrations of chromophores, from which tissue oxygenation levels are derived. Frame averaging can, in principle, be applied at several stages within this pipeline. In practice, the presented compounding algorithm inherently introduces averaging across slices. All PA image

processing was performed using the SIMPA toolkit.

5.2 Validation and results

This section presents the validation of the proposed optical pattern-based framework for 3D PAI (Fig. 5.2), with a focus on the technical contributions critical for clinical translation. Building on the methods described Section 5.1, the evaluation emphasizes performance in key stages of the pipeline: hardware and scan quality, peak detection enhancement via background correction, as well as a validation of temporal volume registration, a post-processing component designed to enable longitudinal PAI studies by correcting for temporal inter-scan misalignment. The section concludes with qualitative benchmarking and a showcase of the optical pattern framework.

The validation is based on the PAD follow-up dataset (Part III Sec. 9), comprising scans from 41 individuals. Each subject underwent four 2D and four 3D photoacoustic acquisitions of the calf muscle before and after a defined exercise protocol. Throughout the scan session, the trident-pattern film remained affixed to the skin to maintain consistent spatial reference across acquisitions. The scanning routine followed a standardized protocol: initial 2D scan, pattern placement and subsequent 2D scan, two consecutive 3D scans, athletic exercise (heel raises), post-exercise 3D scan, post-exercise 2D scan, final 3D scan, pattern removal, and final 2D scan (Fig. 9.1).

All scans were performed by the same operator to reduce inter-user variability and reflect realistic but controlled clinical conditions.

5.2.1 Optical pattern validation

We begin the validation by evaluating the hardware components of the optical pattern-based 3D PAI framework, with particular attention to the technical and clinical requirements defined in Sec. 5.1.1.1. This includes the characterization of ink formulations and substrate materials, as well as a systematic analysis of scan quality under practical conditions.

5.2.1.1 Validation of technical requirements As detailed in the methods section (Tab. 5.1), two iterations of the Macrolex®-based optical pattern were implemented: i) An inkjet-printed version used in the PAD pilot study (Part III Sec. 9), and ii) an improved screen-printed version, where a dielectric was added to increase durability and scalability, deployed in both the PAD follow-up study.

To characterize the optical properties of the patterns, representative samples from each version were placed in a water bath. A multispectral acquisition preset spanning wavelengths from 660 nm to 1200 nm in 10 nm increments was used to acquire a high-resolution spectral response. This setup enabled direct comparison of absorption behavior across the relevant near-infrared window and assessment of pattern visibility at clinically relevant wavelengths.

Inkjet iteration The first Macrolex® prototype (Fig. 5.10) demonstrated the desired optical characteristics for optical pattern-based 3D PAI. Specifically, it produced a strong and clearly visible signal below 700 nm, while appearing transparent at longer wavelengths. In this context, "transparent" was operationally defined as (1) no visually detectable signal in multispectral PAI images, and (2) spectral profiles indistinguishable from background levels, which in this experimental setup were dominated by ringing artifacts and thermal noise. Despite this favorable spectral profile, one notable limitation was observed: the foil substrate itself exhibited a pronounced photoacoustic signal at 1130 nm, violating the design requirement of optical transparency.

In terms of handling, the first prototype performed adequately under controlled conditions. Although the Macrolex® ink was resistant to water exposure (e.g., ultrasound gel) and mechanical rubbing, it was not smudge-proof when touched directly with bare hands. This was likely due to the solubilizing effect of skin oils or sweat. As a result, strict handling protocols were necessary: gloves were required, and direct skin contact with the printed ink was avoided (printed side facing upwards). Under these precautions, the prototype was successfully employed in the PAD pilot study. Nonetheless, it became evident that further improvements in robustness were required for broader clinical applicability.

Screen printing iteration The second Macrolex® iteration (Fig. 5.11) addressed these limitations by reformulating the ink for compatibility with screen printing, a scalable and industry-standard technique. A dielectric was incorporated to enhance adhesion between the ink and the flexible film substrate. This modification significantly improved smudge resistance, even under repeated skin contact, while preserving the desirable spectral properties observed in the first version. Additionally, the screen-printed layer was thinner, potentially reducing unknown ink-induced artifacts.

One remaining limitation was noted: exposure to disinfectants caused partial dissolution of the ink. However, given that the pattern is applied only for short durations (typically $<$1 hour) on intact skin and that the screen-printing process includes a ther-

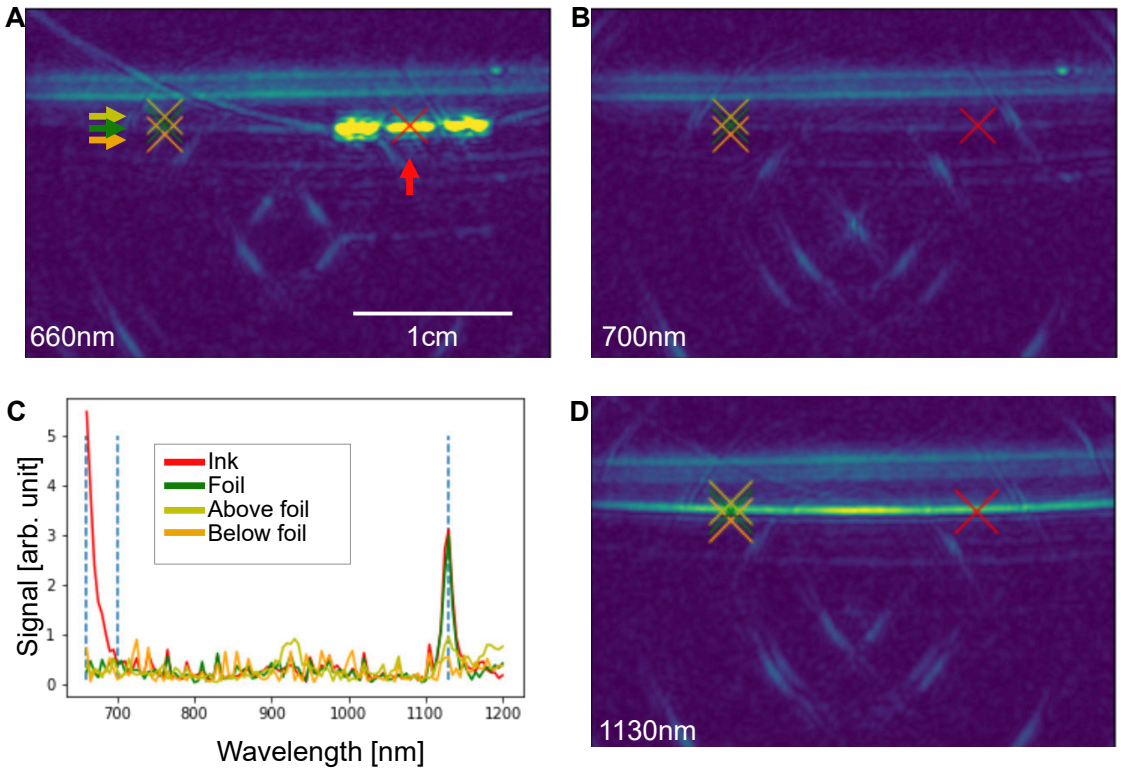


Figure 5.10: Spectral evaluation of the optical pattern (ink and foil) in a water bath using photoacoustic (PA) imaging. **A** PA image acquired at 660 nm, showing strong signal from the pattern ink (bright yellow dots), membrane signals (pale yellow/green lines), and typical artifacts including streaks from high optical absorption and electrical ringing noise of the pre-CE MSOT system. **B** PA image at 700 nm, where additionally weak membrane reflections are visible, primarily due to acoustic impedance mismatch between the water and foil. Pattern signal is absent, indicating transparency of the ink at this wavelength. **C** Spectral profiles acquired from four selected pixels, sampled across the wavelength range from 660 nm to 1200 nm in 10 nm steps. Pixel locations are indicated by matching colored crosses and arrows in the image panels. The ink exhibits strong absorption below 700 nm and transitions to high optical transparency above this threshold. Blue vertical lines indicate the specific wavelengths corresponding to panels **A**, **B**, and **D**. **D** PA image at 1130 nm, where a strong signal is generated by the foil substrate itself (bright yellow), consistent with the spectral peak observed in panel **C**.

mal post-treatment step (used to evaporate residual solvent and akin to sterilization practices), the final device is, to the best of my knowledge and belief, safe for short-term clinical use.

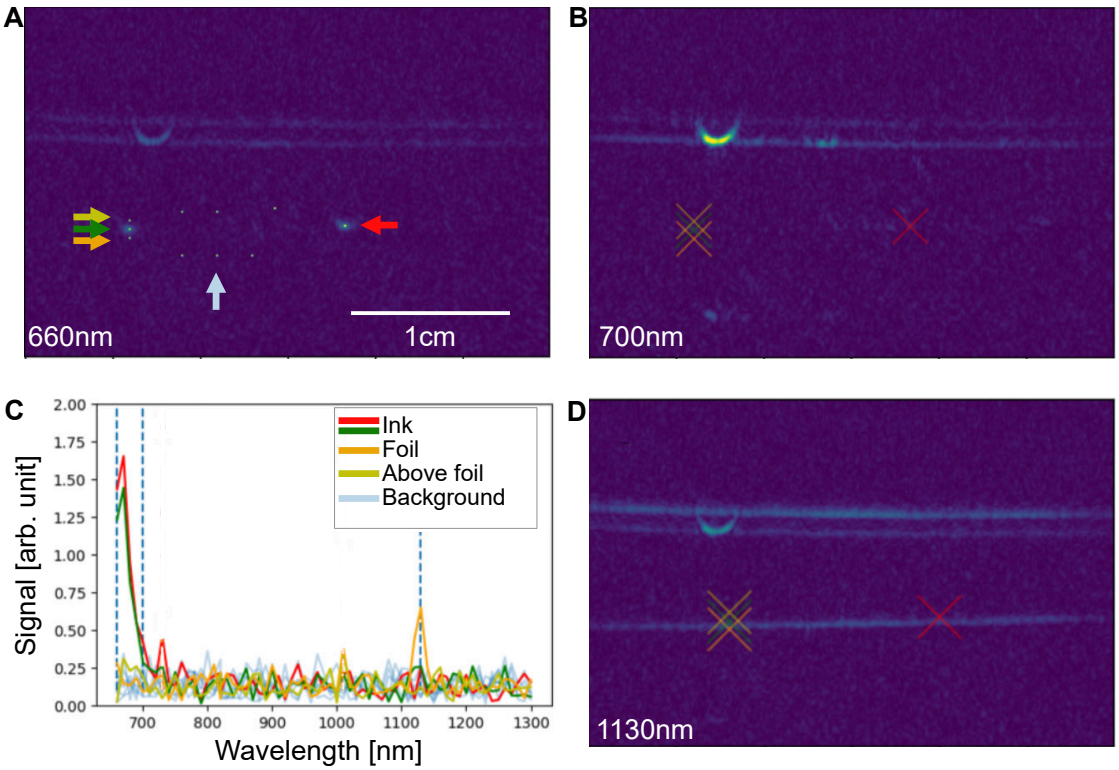


Figure 5.11: Spectral evaluation of the second-generation screen-printed optical pattern for photoacoustic (PA) imaging. Panels replicate the wavelength-specific analysis shown in Fig. 5.10, with PA images acquired with the CE MSOT system at 660 nm, 700 nm, and 1130 nm illustrating key optical responses: **A** At 660 nm, the pattern ink is clearly visible. **B** At 700 nm, the pattern is no longer detectable, confirming optical transparency. **D** At 1130 nm, a strong signal from the foil substrate is observed. **C** Spectral curves plotted in the right panel correspond to specific pixel locations marked by colored crosses, dots, and arrows in the PA images, confirming that the screen-printed pattern maintains strong absorption below 700 nm and high transparency at longer wavelengths, consistent with the first prototype.

Influence on spectral unmixing To validate that the pattern ink does not interfere with spectral unmixing, particularly the estimation of sO_2, beyond controlled laboratory conditions, an in vivo assessment was conducted. While previous spectral measurements confirmed that the ink exhibits no detectable photoacoustic signal above 700 nm, in vivo imaging may be affected by additional factors such as acoustic reflections or unknown factors.

To investigate these potential influences, the PAD follow-up dataset was used, specifically the pre-exercise 2D scans acquired with and without the optical pattern attached.

These semi-paired scans of the human calf allow direct comparison of sO_2 estimations under both conditions. Although the scans are not spatially identical, due to the physical attachment of the pattern between acquisitions, a degree of anatomical overlap is maintained, enabling a reasonable comparison within regions of similar tissue type.

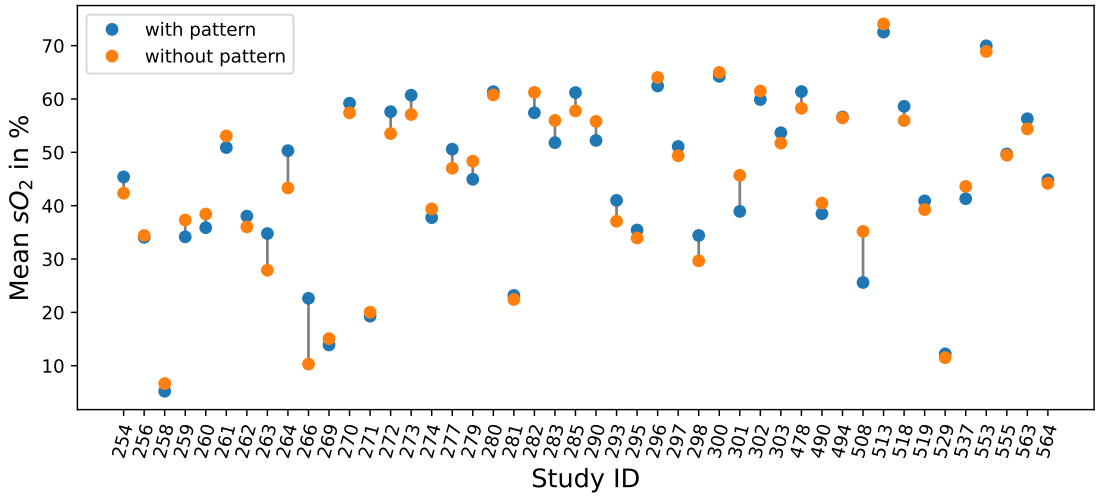


Figure 5.12: Comparison of mean oxygen saturation (sO_2) values per individual (study ID), measured with (blue) and without (orange) the optical pattern. Each circle represents the mean sO_2 value of a single scan, derived from a standardized region of interest (ROI) measuring 20 mm × 5 mm. The average difference in sO_2 between measurements with and without the pattern is 0.5%, with a standard deviation of 3.7%, suggesting a minimal impact of the pattern on oxygenation estimation.

By calculating the mean difference in sO_2 estimates across corresponding regions with and without pattern (Fig. 5.12), the influence of the pattern on functional signal interpretation was quantified (mean ΔsO_2 0.5% $\pm$ 3.7%). As expected, minor differences attributable to physiological variability and scan alignment were present. However, no systematic bias or regionally elevated artifacts were observed in the sO_2 maps derived from patterned acquisitions.

Within the US images an overall drop in mean and standard deviation of the measured signal was observed (Appendix Figs. A1 & A2).

5.2.1.2 Validation of clinical requirements and constraints Since it is not possible to sharply separate clinical requirements from the derived constraints, the following section provides a sequential analysis of biocompatibility, scan quality, and an assessment

of the flat surface constraint.

Biocompatibility The International Organization for Standardization (ISO) 10993-1 standard provides the current best practice for biocompatibility testing. Based on the intended use, application on intact skin for less than 24 hours, a biological evaluation plan was commissioned through TÜV Süd. At the time of thesis submission, results from the biosafety analysis are not yet available.

Scan quality assessment For the following analysis, all volumetric scans from the PAD follow-up dataset (Part III Sec. 9) acquired using the trident optical pattern and an eight-wavelength preset [665, 730, 760, 800, 665, 850, 930, 1030] nm were evaluated.

Scan speed The minimal out-of-plane resolution (equivalent to the slice thickness) of the PAI transducer used in this study is approximately 1.2–2.0 mm, reported by previous performance characterizations [41]. To ensure adequate spatial sampling in the reconstructed 3D volume, the voxel spacing in the scan direction (z-axis) was set to 1.0 mm. This follows the Nyquist criterion requiring "at least two pixels per resolution element" [67].

To satisfy this condition, the transducer must not advance more than 1.0 mm during one full cycle through all wavelengths in the preset. For an 8-wavelength preset acquired at 25 Hz, this constraint translates into a maximum permissible scan speed $v_{z_{\max}}$ with the unit mm per sequence [mm/seq] of:

$$\begin{aligned} v_{z_{\max}} = 1 \cdot \frac{\text{mm}}{\text{seq}} &= \frac{1\,\text{mm}}{\text{time to acquire 8 wavelengths}} \\ &= \frac{1\,\text{mm}}{8/25\,\text{Hz}} = \frac{25\,\text{mm}}{8\,\text{s}} \\ &= 3.125\,\frac{\text{mm}}{\text{s}}. \end{aligned} \tag{5.9}$$

Across the PAD follow-up dataset, the observed median scan speed was 0.41 mm/seq (Fig. 5.13), well below the theoretical upper limit. This oversampling ensures that each voxel contains multiple slices, thereby enabling frame averaging, which in turn improves the signal-to-noise ratio (SNR).

Angular displacement However, the effective displacement of image pixels also depends on the angular rotation of the transducer about the y-axis during scanning.

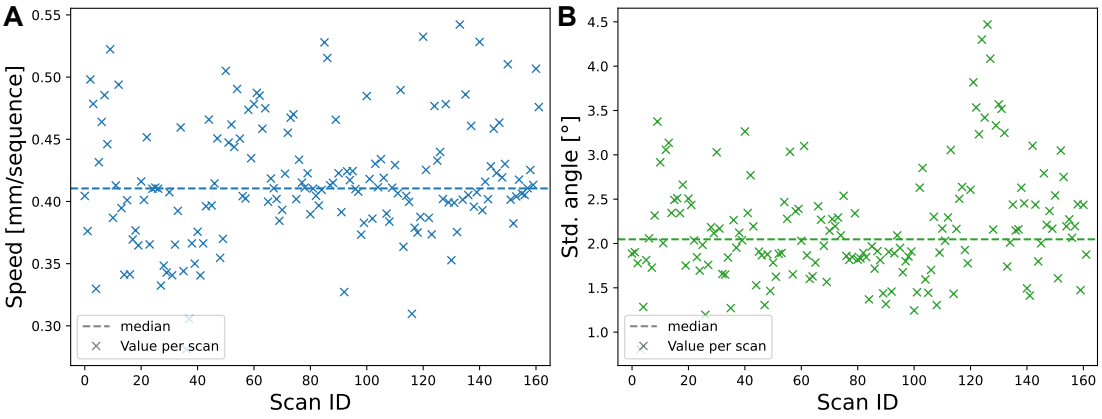


Figure 5.13: **A** Scan speed in millimeters per sequence (mm/seq) across all 3D scans with the trident optical pattern in the peripheral artery disease (PAD) follow-up study, with a median speed of 0.41 mm/seq. **B** Intra-scan rotational stability, expressed as the variance of the rotation angle α, with a median angular variation of 2.05°.

This introduces an additional velocity component (Fig. 5.5), which scales with both the angular deviation β and the distance from the rotational axis d_r. Using a realistic rotation estimate of $\beta = 2°$ and a maximal radial distance of $d_r = 15$ mm yields a displacement $\delta_{\alpha z}$ per sequence of approximately:

$$\sin(\beta) = \frac{\delta_{\alpha z}}{d_r} \quad \Rightarrow \quad \delta_{\alpha z} = \sin(\beta) \cdot d_r \approx 0.5 \text{ mm}. \tag{5.10}$$

Even with this additional displacement, the total effective scan velocity on average remains within the acceptable bounds for spatial sampling (0.5 mm/seq + 0.41 mm/seq $\leq$ 1 mm/seq), affirming the robustness of the acquisition protocol.

Pattern visibility To assess the practical visibility of the pattern throughout each scan, the effective scan length in the z-direction was used as a proxy. This approach is justified by the high detectability of the pattern peaks in the raw 2D PA images. In more than 98% of axial slices across the dataset, all peaks of the trident pattern were successfully detected (Fig. 5.16).

For the 164 volumetric scans analyzed, the median scan length in the z-direction was 41 mm (Fig. 5.14). Given the physical dimensions of the trident pattern used (30 mm in width and 50 mm in length) with a minimum resolvable distance to the origin of $d_{\text{min}} = 1.5$ mm (Tab. 5.2), the theoretical maximum scan length s_{max} is 45 mm, as defined by Eq. 5.11:

$$s_{\max} = p_{\text{length}} - \left(\frac{p_{\text{length}}}{p_{\text{width}}/2} \times d_{\min} \right). \tag{5.11}$$

This relationship is derived from the trigonometric geometry of the trident pattern design, where the opening angle γ governs the FOV and is given by:

$$\tan(\gamma) = \frac{p_{\text{width}}/2}{p_{\text{length}}} = \frac{d_{\min}}{p_{\text{length}} - s_{\max}}. \tag{5.12}$$

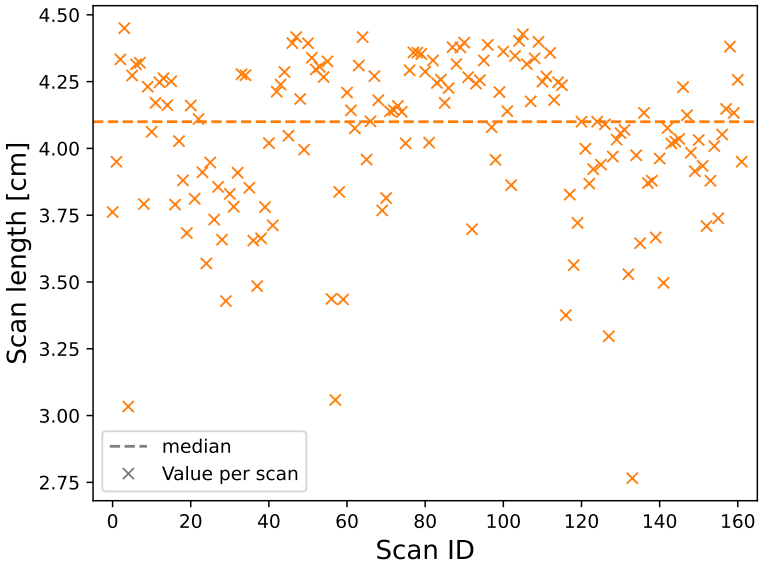


Figure 5.14: Scan length in the scan (z-) direction across all 3D scans with the trident optical pattern in the peripheral artery disease (PAD) follow-up study, shown in centimeters. The median scan length is 4.1 cm.

Based on this analysis, the achieved median scan length corresponds to more than 91% of the theoretical maximum 3D FOV. This demonstrates that the clinician was able to reliably capture the majority of the available volume during routine imaging, even under real-world clinical conditions. The observed consistency in scan length across the dataset further supports the pattern's robustness and usability in volumetric clinical acquisitions.

Flat surface constraint "To analyze the potential effect of violations of the flat surface assumption, we selected a PA image [...] that featured a particularly high curvature of the [foil]. We then approximated the geodesic distance between the [pattern] points in one image slice with a manually placed Bézier curve using two control points, as shown

in [Appendix] Fig. A3. The resulting increase in distance compared to the Euclidean baseline was 2% (17.7 mm vs. 17.4 mm). This discrepancy corresponds to a shift of $a_0 = 0.4$ mm in the pattern coordinate system and an angular error of $\alpha = 1.7°$. This can be regarded as tolerable given the overall resolution of the imaging system and the fact that we picked one of the most extreme cases" [140] and since the magnitude of the shift introduced by this effect is approximately equal to the probe movement between two consecutive frames. The data supporting this investigation is derived from results previously published in [140], where photoacoustic volumes of the human forearm were acquired in a registered healthy volunteer study (German Clinical Trials Register: DRKS00023205). In that study, the initial high-absorbing ink prototype of the optical pattern was utilized.

5.2.2 Validation of the 3D reconstruction algorithm

Following the validation of the hardware and requirements of the optical pattern, the subsequent analysis focuses on the algorithmic core of the framework: pose estimation and spatial compounding, the slice-to-volume distribution and filling step (Part I, Sec. 2.2.3). This section begins with the validation of the detection algorithm, including the effect of background correction on peak visibility. This is followed by a brief assessment of pose estimation performance and, consequently, the overall reconstruction accuracy. Subsequently, the performance of the revised interpolation method is qualitatively analyzed and benchmarked. The section concludes with an evaluation of the temporal volume registration.

Components of the pipeline that have not undergone significant modification such as pattern design, line registration, coordinate computation, and signal filtering are not re-evaluated in detail. For a comprehensive analysis of their performance and error propagation, the reader is referred to the prior validation study in [185], which provides a thorough examination of the geometric and computational reliability of the original optical pattern approach.

5.2.2.1 Validation of the detection of the optical pattern

Background correction As outlined in the methods section, pose estimation relies on a sequence of pre-processing steps to isolate the pattern peaks from background structures. The introduction of a less absorbent ink selected for improved clinical robustness necessitated an additional correction step to enhance peak visibility.

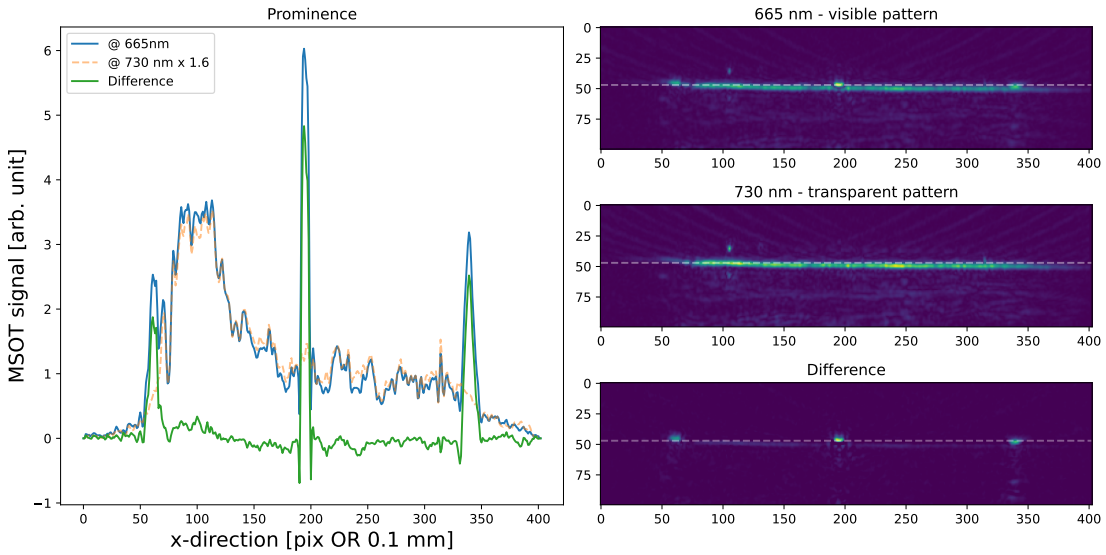


Figure 5.15: Background correction for improved peak detection in photoacoustic (PA) images of the optical pattern. **A** Line plot of PA signal intensity across the height of the pattern at 665 nm (blue) and 730 nm (dashed orange, scaled by a factor of 1.6), along with their difference (green), representing the result of the background correction. (**B**–**D**) Corresponding 2D PA images at 665 nm, 730 nm, and the corrected image, respectively. The left peak of the pattern becomes clearly distinguishable from the background only after applying the correction, emphasizing the benefit of background subtraction for reliable peak detection.

Figure 5.15 illustrates the qualitative impact of background correction, using the 730 nm channel as correction wavelength. The line profile shown in Figure 5.15 **A** reveals that, after correction, the pattern peaks stand out with intensity levels several orders of magnitude higher than those associated with background structures such as skin or soft tissue features.

Appendix Fig. A4 further demonstrates the wavelength dependency of the background correction. When using a suboptimal wavelength with markedly different tissue absorption characteristics such as 930 nm the background correction becomes ineffective. In contrast, a broad range of wavelengths between 730 nm and 850 nm have been shown to perform reliably for background correction, offering both flexibility and robustness across different imaging conditions. To quantify the benefit of the background correction, the following analysis evaluates the detection rate of pattern peaks across the PAD follow-up dataset.

Detection The consistency of peak detection is critical for reliable pose estimation in the optical pattern-based 3D reconstruction pipeline. To quantify performance, the detection rate is defined as:

$$r_{\text{det}} = \frac{\text{detected frames}}{\text{expected frames}} = \frac{n_{\text{frames}}}{n_{w,\text{pat}} \cdot \left(\frac{f_{\text{last}} - f_{\text{first}}}{n_{w,\text{preset}}} \right)}, \tag{5.13}$$

with:

- r_{det}: detection rate;
- f_{first}: index of the first frame of the 2D image stack with detected coordinates;
- f_{last}: index of the last frame with detected coordinates;
- $n_{w,\text{ pat}}$: number of wavelengths below 700 nm;
- $n_{w,\text{ preset}}$: number of wavelengths within the preset;
- n_{frames}: total number of detected frames.

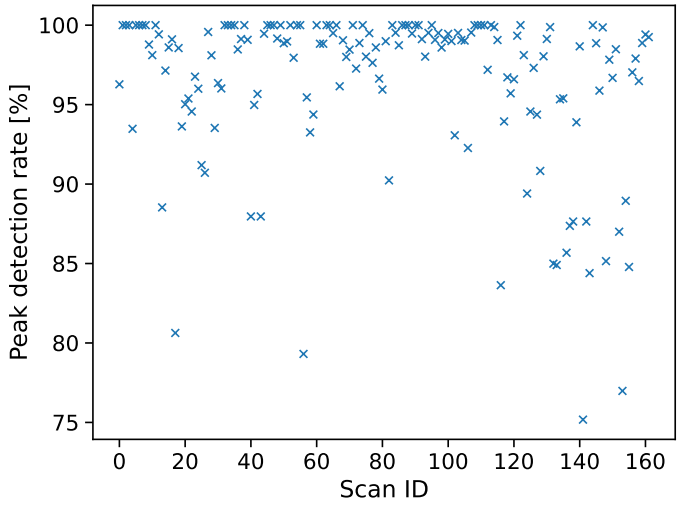


Figure 5.16: Peak detection rate, as defined in Equation 5.13, for all 3D scans acquired in the peripheral artery disease (PAD) follow-up study with the optical pattern trident design. Each value represents the proportion of frames in which all three peaks of the optical pattern were successfully detected. The results demonstrate a high robustness of the pattern-based pose estimation, with a median detection rate of 98.7% across the dataset.

A slice was considered *detected* only if all three peaks of the trident pattern were identified. Frames in which only one or two peaks were labeled are defined as *undetected* since full coordinate calculation requires a complete peak set. Across the PAD follow-up dataset, a median of 98% of slices per scan were successfully detected (Fig. 5.16).

Following initial detection, a Gaussian fit was applied to refine the peak positions. This post-processing step smooths the detected peak transitions across frames and helps prevent false discontinuities in the reconstructed 3D volume. Figure 5.17 qualitatively illustrates the improvement in reconstruction fidelity using Gaussian refinement on a PA image with the iThera provided beamforming (DAS, no envelope detection). In particular, the resulting pattern lines appear visibly smoother, and the effective FOV defined by the continuous detection of peaks was expanded. In the illustrated example, the crop region (highlighted by a yellow dashed line) is significantly larger when Gaussian fitting is used. Without this refinement, the frame-to-frame distance exceeded the `peak_thres` parameter (Tab. 5.2), causing the limited FOV.

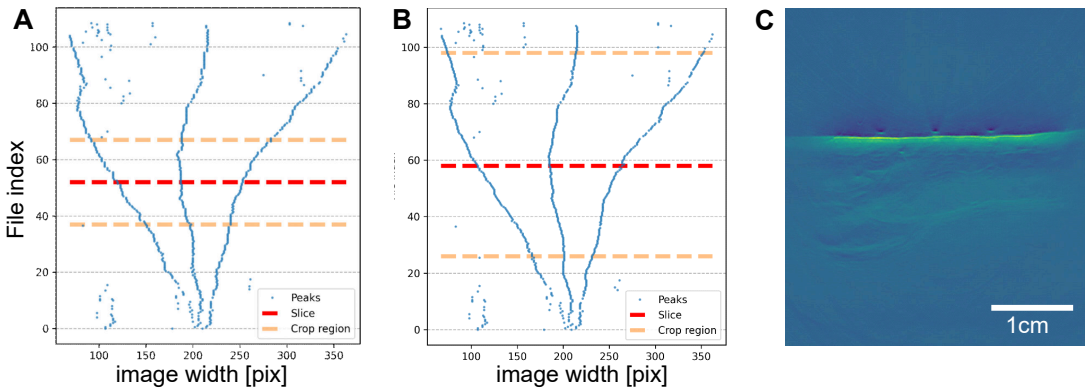


Figure 5.17: Effect of Gaussian refinement on peak detection and reconstructed field of view (FOV) in a 3D photoacoustic scan using the optical pattern. **A** Bird's-eye view (x–z plane) of detected pattern peaks across all frames without Gaussian smoothing. **B** Corresponding view after applying Gaussian smoothing to refine peak positions. The yellow dashed lines denote the crop region used for volumetric reconstruction, while the red line marks the center axis of the FOV. Gaussian refinement results in a more continuous set of detected peaks and an expanded usable scan range defined by `peak_thres` (Tab. 5.2). **C** Representative photoacoustic image slice (e.g., corresponding to the dashed red line in A) reconstructed using the iThera beamforming algorithm before applying background correction.

5.2.2.2 Pose estimation The performance of the pose estimation method was rigorously evaluated in the context of my master's thesis [185] and further summarized in the subsequent publication [140]. Specifically, the accuracy of the registration was quantified using an N-wire phantom, where the proposed pose estimation approach achieved a mean registration error of 0.63 mm. This performance was superior to that obtained with an optical tracking system, which exhibited a registration error of 0.87 mm. Both measurements were derived using an iterative closest point algorithm.

To assess the reproducibility of the method in a clinical setting, an in vivo experiment was conducted involving seven measurements of a blood vessel in two human volunteers. The mean distance between corresponding vessel measurements across repeated acquisitions was 0.63 mm in one volunteer and 0.80 mm in the other. These results demonstrated good consistency, particularly when considering that the tracking reference system used in the experiments exhibited a positional error exceeding 1 mm. Further reading on methodological details can be found in [140].

5.2.2.3 Compounding To validate the new interpolation strategy used for hole filling during volume compounding, a two-step evaluation was conducted. The first step verified that the newly implemented method produced equivalent interpolated values compared to the initial reference implementation based on `scipy.griddata` with linear interpolation. The second step involved a qualitative benchmarking and performance analysis to optimize runtime efficiency while maintaining reconstruction quality.

Figure 5.18 compares four interpolation results: the original volume with holes (i.e., unfilled voxels), the reference `scipy.griddata` linear interpolation, an initial custom implementation with hole identification performed in a separate step, and a fully optimized custom interpolation with integrated hole detection as described in the methods section. The assessment was performed on two datasets: a synthetic "toy" image with gradually varying pixel values and artificial holes, and an in vivo PAI volume of a human calf, showing a vascular tree.

The pixel values of the custom implementation matched those of `scipy.griddata` in all regions except: (1) volume edges, where the griddata implementation performs slight extrapolation while the initial custom method discards edge slices with more than 70% missing voxels, and (2) elongated holes spanning more than one voxel in the z-direction, which griddata fills but the custom method leaves unfilled by design. These differences are intentional and align with the goal of avoiding unreliable interpolations in underconstrained regions.

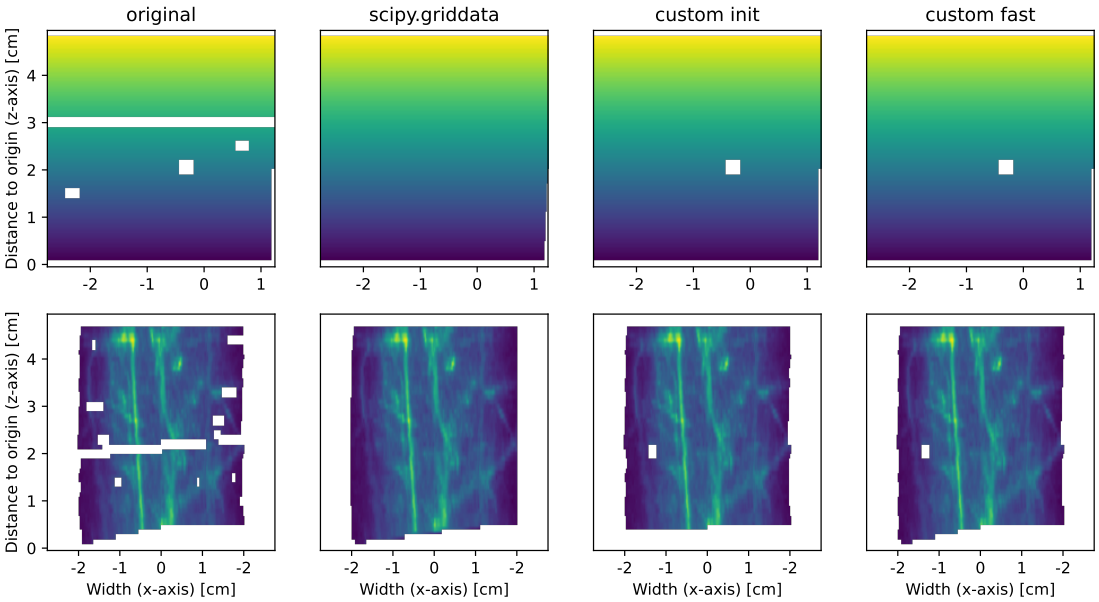


Figure 5.18: Comparison of interpolation strategies applied to 3D photoacoustic volumes with missing voxel regions. From left to right: the original input without interpolation, linear interpolation using `scipy.griddata`, an initial custom method with separate hole identification, and an optimized custom implementation with integrated interpolation logic. Top row: synthetic "toy" dataset with gradually varying intensity values and artificially introduced holes. Bottom row: in vivo volume showing a representative vascular structure. All images are x–z plane intensity projections. White patches indicate unfilled voxels. Notable differences between methods appear at volume edges where griddata extrapolates, the initial custom method crops, and in regions with contiguous missing voxels along the z-direction, which are only filled by griddata.

Runtime performance To assess runtime performance, Figure 5.19 reports average interpolation times across 100 real in vivo volumes. Each optimization step in the custom implementation yielded a relevant speedup. All calculations for benchmarking were performed using central processing unit (CPU) processing on a Microsoft Windows 10 (11 for Figs. 5.19, A6) Enterprise, 64 GB random access memory (RAM), AMD Ryzen 9 5900X (12 cores at 3.7 GHz) machine.

5.2.2.4 Assessment of the temporal volume registration We now turn to a key post-processing step: refined, image-based registration for spatial alignment of PAI volumes acquired over time. As in the previous sections, the PAD follow-up dataset (Part III

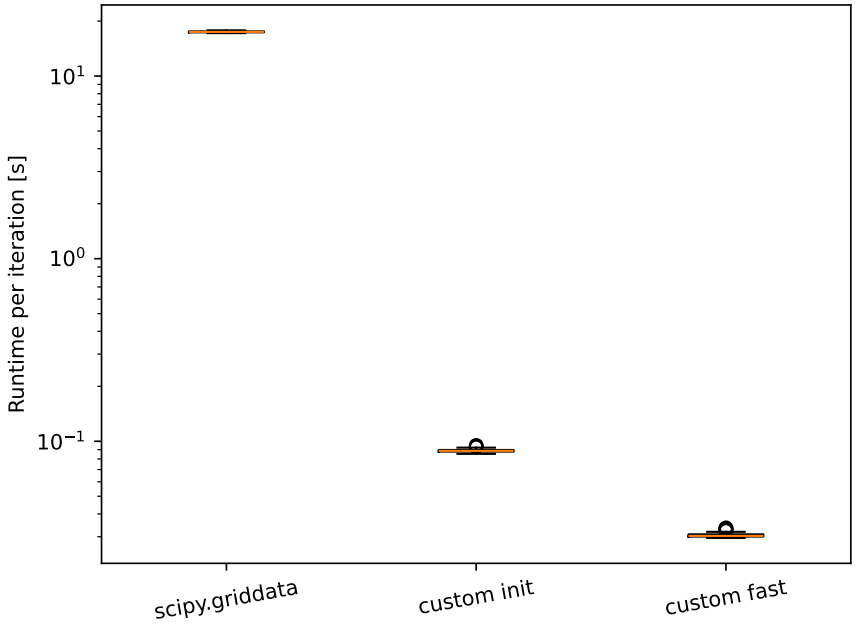


Figure 5.19: Qualitative benchmarking of 100 interpolation runs on real 3D photoacoustic volumes, corresponding to the in vivo dataset shown in Fig. 5.18. Each successive improvement in the custom interpolation implementation resulted relevant speedup in average runtime (scipy.griddata 17.47 ± 0.14s; custom init 88.8 ± 1.6ms; custom fast 30.6 ± 0.9ms). Benchmarking results on synthetic toy data are provided in the Appendix (Fig. A6).

Sec. 9) was used as an in vivo example. This dataset consists of four volumetric scans of the human calf, acquired over a short time interval (<1h). During this period, the optical pattern remained affixed to the subject, ensuring consistency in the reference frame across scans. The first scan, acquired with the subject in a resting, prone position, was defined as the fixed image for the registration process. The second scan, referred to as the *pre* scan, was acquired immediately thereafter. The remaining two *post* scans were recorded after a brief athletic exercise session involving heel raises. This required the subject to stand, perform physical activity, and then return to a lying position for subsequent scans. Following the rigid registration of two example volumes with extreme displacement, a clear visual and quantitative improvement in anatomical alignment was observed across time points (Fig. 5.20). The clinical application of the temporal location matching for PAD classification is presented in Section 9.2.3.

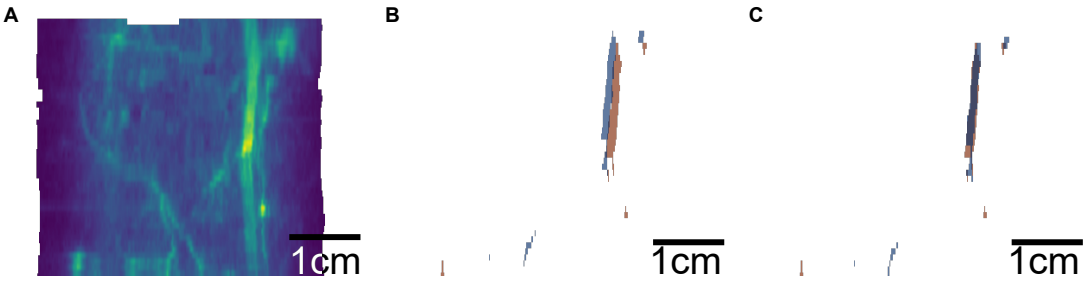


Figure 5.20: Rigid registration of temporally spaced in vivo 3D photoacoustic imaging (PAI) volumes of a human calf. **A** x-z projection of the fixed scan showing prominent superficial vessel structures (yellow), located below the skin surface. **B** Pre-registration overlay of segmented vessels from the fixed (orange) and moving (blue) scans, highlighting spatial displacements in the x- and z-directions (threshold 98th signal quantile - 82th quantile Appendix Fig. A5). The superficial vascular tree serves as anatomical landmarks for landmark-based registration (Fig. 5.6). **C** Post-registration result showing improved alignment of vascular structures, confirming successful spatial matching of the temporally separated volumes. The Dice similarity coefficient increased from 0.47 to 0.66, indicating enhanced registration accuracy.

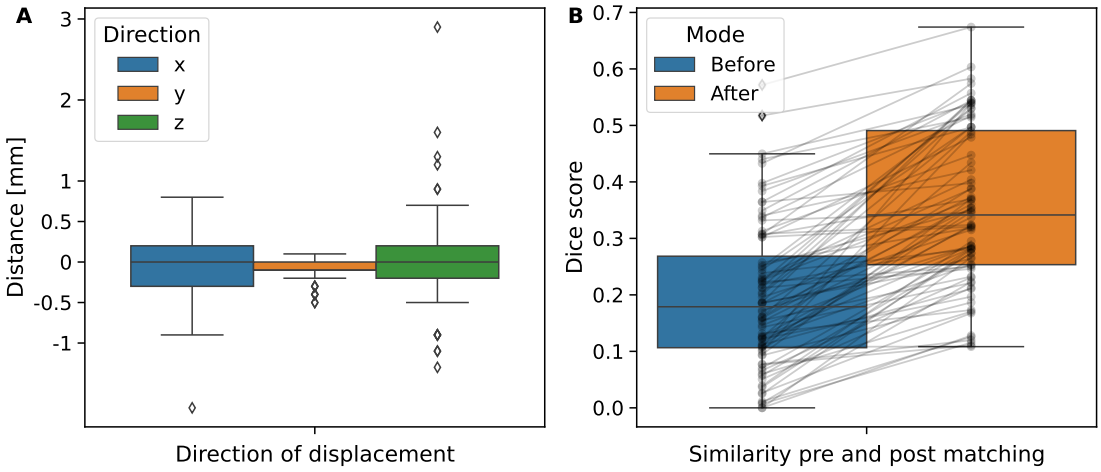


Figure 5.21: Quantitative evaluation of temporal volume registration in an in vivo photoacoustic imaging (PAI) dataset of the peripheral artery disease (PAD) study (Part III Sec. 9). **A** Spatial displacement between temporally acquired 3D PAI volumes after initial pattern-based registration and prior to registration refinement, expressed in millimeters. Displacements are calculated relative to the fixed baseline scan and visualized per direction. These shifts represent inter-scan variability. **B** Registration performance assessed using intensity-thresholded vessel segmentations. The Dice similarity coefficient (DSC) between fixed an moving scan is shown before and after rigid registration.

Quantitative assessment revealed that 27% of the volumes exhibited spatial offsets exceeding one voxel spacing between the fixed and corresponding moving image, an operational threshold used to define displacement. The most frequent displacements occurred in the x-direction, while the largest absolute shifts were observed in the z-direction (the scanning axis), likely due to minor pattern displacement across the tissue surface during acquisition (Fig. 5.21 **A**). Further evaluation using the DSC confirmed a quantitative improvement in spatial congruency following registration (Fig. 5.21 **B**).

5.2.3 Implementation assessment

Showcase of the optical pattern framework A system-level test was performed to demonstrate the integrated functionality of the complete optical pattern-based 3D PAI processing pipeline.

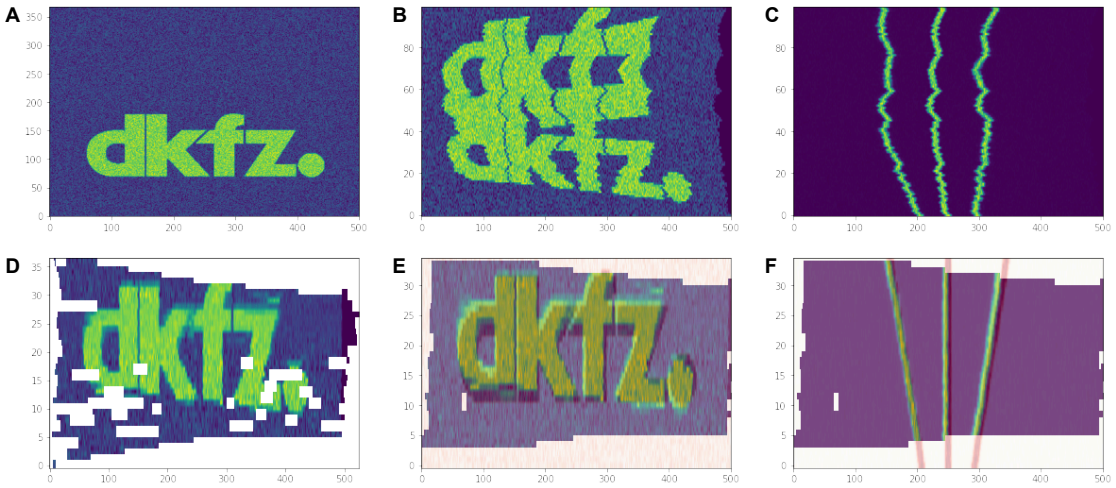


Figure 5.22: System-level test of the complete optical pattern-based 3D photoacoustic imaging (PAI) processing framework using a synthetic volume with **A** an embedded *dkfz.* logo and pattern. Input slices sampled from the synthetic volume at varying angles, including forward and backward motion, at the **B** logo and **C** pattern level. The corresponding reconstruction results are shown after applying the full pipeline, including pose estimation, compounding **D**, and interpolation at **E** logo and at **F** pattern level. While the x-axis is always in [pix], the y-axis is in [pix] **A**, [frames] **B** and **C**, and in [mm] **D**-**F**.

As shown in Fig. 5.22, an artificial volume (3D array with Gaussian noise) was created with a *dkfz.* logo embedded as structural content in depth and with a synthetic optical pattern on top. A series of 2D slices was sampled from the volume, including both

forward and backward probe motion, and subsequently reconstructed using the full framework. This test serves as an end-to-end validation of the entire processing chain and enables the identification of potential failure points resulting from complex pipeline interactions. The accurate reconstruction of a recognizable logo across non-uniform trajectories confirms both the spatial consistency and robustness of the system.

Qualitative framework benchmarking In addition, a qualitative benchmarking comparison was conducted between the original pattern-based prototype pipeline [185] and the improved framework introduced here.

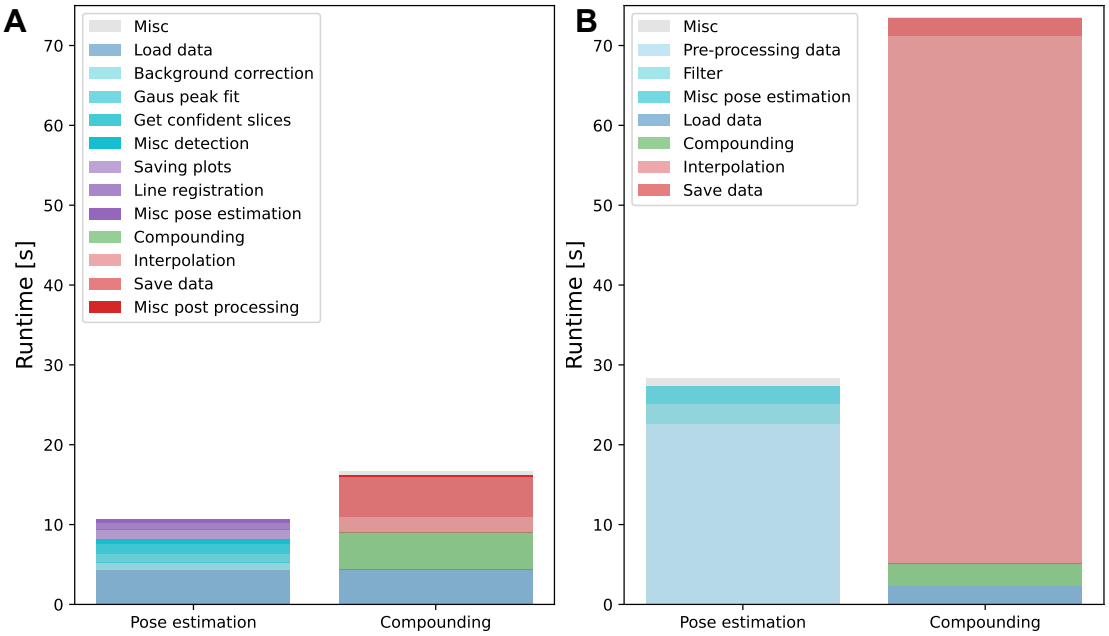


Figure 5.23: Benchmarking comparison between the improved, clinically-oriented optical pattern-based reconstruction pipeline presented in this section **A** and the original **B**. The two versions are only partially comparable, as the original implementation lacked several key features now integrated for robustness and quality assurance. An additional background correction step (runtime 0.145 s) was applied to the original pipeline to avoid processing failure. Despite the enhanced functionality and added computational steps, the improved implementation demonstrates a significantly reduced overall runtime, driven primarily by optimized preprocessing and interpolation routines. Benchmarking of processing a complete dataset, including ultrasound and unmixed data, which is only possible using the new framework, is shown in Appendix Figure A7. *Misc* refers to runtime elements that are not directly related to the core functionality of the framework, but instead cover aspects such as module interaction, initialization procedures, and other auxiliary processes.

Despite incorporating several new components such as background correction, pose refinement, and optimized interpolation, the total processing time per scan has been substantially reduced. Specifically, the average processing time decreased from approximately 100 seconds to 27.4 seconds (Fig. 5.19), while maintaining or improving reconstruction quality.

The most significant speedups were achieved in the computationally intensive preprocessing and interpolation stages. While some other steps became slightly more complex to improve accuracy and reproducibility, the overall pipeline now supports processing times well below acquisition time. For the current clinical setup (110 frames, 8 wavelengths, 25 Hz), the acquisition time is approximately 35.2 seconds.

5.3 Discussion

This section presented a comprehensive and robust framework for image-based 3D PAI, leveraging an optical pattern to encode probe pose information directly into the acquired images. By combining hardware innovations with a modular software architecture and a graphical user interface, the system enables volumetric PAI reconstruction accessible also to clinician scientists without a computational background.

The screen-printed optical pattern met all major predefined technical and clinical requirements. As a result, high peak detectability was achieved across all datasets, with detection rates exceeding 98%. This performance was largely enabled by the development of a novel background correction method, which also ensured minimal interference with oxygenation estimates derived from multispectral data.

In addition to enabling accurate pose estimation, the visual guidance provided by the pattern helped clinical operators to achieve a consistent, steady probe movement, which translated into high-quality 3D PAI data. Further optimization of the framework enabled efficient slice-to-volume compounding, including robust distribution and hole-filling strategies. These improvements resulted in processing times shorter than the actual acquisition duration, thereby laying the foundation clinical application. Finally, a temporal volume registration module was introduced as a post-processing step, allowing for the spatial alignment of longitudinal scans, a critical requirement for monitoring disease progression or treatment response over time.

The upcoming interpretation of the results and discussion of limitations is structured into three key areas: Hardware and acquisition, algorithms, and validation strategy.

5.3.1 Hardware and acquisition

"While the general [optical pattern] concept proposed is potentially very powerful, our first prototype implementation comes with several limitations. The main current drawback is the fact that we use a 2D pattern. Future work could be directed to developing a 3D pattern, allowing for the recovery of 6 DoF from a single measurement, thus eliminating the need for holding the probe orthogonally to the pattern. Furthermore, the trident design was chosen for its relatively simple optical pattern. While—in theory—it enables us to uniquely recover the pose of the device from a single measurement and thus fulfills the core requirement, more work needs to be invested in patterns that are optimized for the robustness of pose estimation in the presence of uncertainties [...]. In this context, dyes that absorb at different wavelengths of light may play an important role in the future." [140]

As an alternative to increasing pattern complexity, hardware augmentation with IMUs could offer a pragmatic solution [138]. An IMU could track rotational motion and fan-like movements, providing real-time feedback to the user and potentially improving reconstruction quality by ensuring adherence to motion constraints or by updating the pattern-derived pose. On the downside, an additional IMU would significantly complicate the system design.

Flat surface constraint Another hardware-related consideration is the flat surface constraint, which underpins the current coordinate estimation model. While this assumption simplifies computation, its validity may vary depending on the anatomical site being imaged. Results from the following multimodal fusion experiments (Fig. 6.6 **D**) suggest that deviations from flatness may have limited practical impact in certain soft tissue regions, though this warrants further investigation. Nonetheless, "we see primary ways to address the issues arising from the flat surface constraint in the future: 1) Image processing-based solution: Automatically segmenting the skin from the image would enable us to approximate the geodesic distance along the skin surface between any two given [pattern] points by determining the length of the curved line representing the skin in a 2D tomographic image. As isometric transformations such as the bending of a flexible material (here: the [pattern]) preserve geodesic distances [199], a curved line represented by three [pattern] points can be related to a corresponding straight line in the flat [pattern] coordinate system via the Euclidean distance (which is equal to the geodesic distance on surfaces with zero curvature). Leveraging this principle, the curved [pattern] could be reconstructed in 3D and serve as a reference

coordinate system just like the flat variant. While this approach could compensate for small deviations from the flat surface constraint, a highly curved target anatomy would still pose a challenge when used in conjunction with a 2D pattern. [...] 2) Hardware solution: Alternatively, a semirigid pattern could be applied, comprising a flat [pattern] at the top and a flexible part as a padding layer between the tissue and the rigid part, as illustrated in [Appendix] Fig. A3 (right)." [140]

Ink In terms of the selected hardware components, the ink used for pattern printing Macrolex® Blau 3R was reported in its datasheet to have a transmission of approximately 90% in the 700–740 nm range for a 2 mm layer thickness. Nevertheless, no detectable photoacoustic signal above 700 nm was observed during experiments. This discrepancy could be attributed to several factors: the thinness of the actual printed ink layer, potential reflection losses rather than absorption, or the influence of the vendor's transmission measurement setup, which included a 2% TiO_2 solution, potentially influencing the spectral measurements.

Substrate In addition, the foil substrate used for pattern printing exhibited an absorption peak near 1130 nm. While this wavelength is typically not used for functional imaging of sO_2, it may be relevant for other applications such as spectral unmixing of water, lipids, fat, or collagen, which benefit from longer wavelengths. Fortunately, the absorption peak is spectrally narrow, indicating that selective wavelength exclusion may still enable compatibility with such use cases. Furthermore, while the influence of the pattern on the unmixed sO_2 signal was shown to be minor, the acoustic damping and reflection caused by the foil have been observed to be noticeable. Although this does not substantially affect sO_2 estimation, it may reduce measurement depth and sensitivity to weak signals, particularly in deeper tissue regions. These effects warrant further investigation to fully assess their impact on quantitative accuracy and clinical applicability.

Design The trident pattern design remained unchanged and is identical to that originally introduced and analyzed in [185]. While it has been shown that pose estimation accuracy for this geometry degrades near the pattern origin, overall sub-millimeter accuracy was achievable and appears sufficient for the intended clinical applications. The choice of using exactly three lines in the trident was deliberate: increasing the number of lines not only complicates line matching but also increases sensitivity to localization errors due to reduced inter-line spacing.

During early development, a modified N-shaped pattern inspired by N-wire phantoms commonly used in ultrasound calibration was also investigated and briefly tested in the PAD study (Part III Sec. 9). However, an analytical revision of the pose estimation model revealed that a single N-shape results in pose ambiguity, specifically in the sign of the rotation angle α (Appendix Fig. A8). The geometry permits only an absolute value to be recovered, making the pose estimation non-unique. To resolve this, either a fourth guiding line or a geometric tilt (basically going back to the original trident design) is required to uniquely determine the rotation sign.

5.3.2 Algorithms

A major limitation of the proposed optical pattern-based framework for 3D photoacoustic imaging lies in its current development status: the software was not implemented under formal medical device regulatory standards. As such, clinical translation and large-scale deployment would almost certainly require a complete reimplementation from scratch in accordance with relevant regulatory frameworks, including quality assurance, traceability, and risk management standards.

Despite this limitation, the current prototype has demonstrated the technical feasibility and robustness of the approach. Importantly, the system achieves processing times shorter than acquisition times, which is a crucial milestone for integration into time-sensitive clinical workflows.

Interpolation This performance gain was primarily made possible through multiple software optimizations, most notably, a redesign of the interpolation step used for filling missing voxel values during slice-to-volume reconstruction. While interpolation remains a tradeoff between speed and spatial fidelity, the current implementation balances these needs effectively for most clinical use cases. However, applications requiring higher spatial continuity or image realism might benefit from switching to voxel-based reconstruction, albeit at the cost of significantly increased computational overhead.

Temporal volume registration Another key addition to the software framework is the temporal volume registration module, enabling longitudinal analysis of repeated 3D scans. In this feasibility study, temporal registration focused solely on rigid translational alignment of scans from the same anatomical region, with the optical pattern remaining in place between time points. Non-rigid tissue deformation effects were not explicitly

modeled.

A notable pitfall in the initial implementation of the approach is the use of vascular landmarks for alignment: pressure applied during PAI acquisition can compress venous structures, causing them to disappear in subsequent scans and reducing registration robustness. Moreover, the visibility of superficial vessels varies significantly across anatomical regions and patient populations, which further limits general applicability. To address these limitations, future work may focus on incorporating US data into the registration process. Unlike PAI, US imaging consistently captures the anatomical structure of tissue, making it a more stable reference modality. Consequently, US-based rigid or even non-rigid registration could serve as a more reliable backbone for temporal alignment.

In parallel, a more robust and interpretable metric is needed to assess registration accuracy. The simple threshold-based segmentation used in the proof-of-concept to extract vascular landmarks produces noisy binary masks. The presence of small outlier structures makes contour-based metrics such as the Hausdorff distance overly sensitive and less reflective of meaningful improvements. If vascular structures continue to serve as the basis for registration, more advanced segmentation methods should be explored. These would enable the extraction of vessel centerlines, making it possible to use topology-aware metrics such as the centerline Dice score proposed by Shit et al. [200].

5.3.3 Validation

Validation of the optical pattern-based framework demonstrated high robustness, particularly in peak detection of the new ink, which is critical for successful 3D reconstruction. Comparisons to earlier versions of the framework without enhanced detection routines are not meaningful, as those systems failed to reconstruct volumes from data acquired using the new pattern entirely.

Detection It is important to acknowledge that the definition of a *detected peak* may introduce bias into the detection rate metric, as the total number of expected frames (Eq. 5.13) is computed based on the first and last frame in which a peak was successfully detected. This implies that any undetected frames near the pattern boundaries are excluded from the calculation. While this exclusion is reasonable given that peaks at the edges are often not visible in the PA image due to limited FOV it does introduce a systematic underestimation of undetected frames. However, when this metric is

interpreted in conjunction with scan length, which was shown to be close to the theoretical maximum, it becomes a robust indicator of acquisition consistency and spatial coverage. Nonetheless, it must be noted that the reported detection rate may be slightly overconfident.

Acquisition During image acquisition, the optical pattern could be identified visually in the 2D MSOT live view. However, this was not always straightforward: sometimes only a central peak was visible. In these cases, clinicians had to rely on experience and trust that the post-processing routines, particularly background correction, would recover the full pattern signal retrospectively. When paired with the scan speed and angular deviation analyses presented earlier, the results suggest that while the optical pattern does provide visual guidance during acquisition, its effective use requires operator training.

Benchmarking The benchmarking of the framework's performance must also be interpreted with care. Runtime is highly hardware-dependent, particularly sensitive to variables like disk speed and RAM. Therefore, performance analysis was conducted qualitatively, using a representative scan to establish a realistic order of magnitude. Due to the lack of a direct API interface with the MSOT system, end-to-end pipeline optimization was not the primary focus of this work. Instead, efforts concentrated on demonstrating the feasibility, modularity, and robustness of the framework in both controlled and clinical settings.

5.3.4 Comparison to related work

As summarized in the related work (Part I Sec. 3.3), various strategies have been explored to extend PAI from two- to three-dimensional imaging. These methods generally balance image quality, system complexity, and clinical practicality. Motorized or mechanically actuated scanners provide high spatial precision and dense sampling but at the cost of bulky hardware [118]. Optical or electromagnetic tracking-based approaches [133, 135] enable flexible handheld operation but introduce calibration demands and hardware dependencies that complicate routine use. In contrast, the optical pattern approach developed in this work integrates spatial encoding directly into the acquisition, avoiding external tracking systems and maintaining compatibility with standard probes. It thus offers a low-cost, device-independent, and easily deployable alternative for volumetric imaging. While highly specialized systems such as breast

scanners [124] achieve excellent image quality in dedicated use cases, their limited flexibility hinders broader clinical translation. Recent sensorless or learning-based compounding methods [146] promise similar benefits without added hardware but remain prone to drift and instability under real-world conditions. Overall, the optical pattern framework presented here provides a practical middle ground, combining robustness and simplicity with sufficient volumetric precision to support future clinical applications.

5.3.5 Conclusion

Overall, the introduced improvements substantially mitigated two of the primary limitations of the optical pattern-based approach to 3D photoacoustic imaging: the requirement to detect discrete image features (i.e., peaks) and the presence of holes in the reconstructed volume caused by freehand scanning variability. These developments place the method on a more competitive footing compared to mechanical 3D scanning approaches that do not inherently face these issues, such as the system proposed by Lee et al. [118] while preserving key advantages including device independence, application flexibility, and low-cost implementation.

In addition, the framework inherently supports temporal location matching, which is critical for longitudinal clinical studies. This feature is particularly relevant as volumetric PAI continues to move toward dynamic, multi-time-point monitoring of functional biomarkers. This aligns with recent work by Lefebvre et al. [201], who evaluated image co-registration strategies in photoacoustic mesoscopy. Their findings showed promising results of intensity-based registration methods. However, deep learning-based approaches seem to offer superior performance, especially in preserving the continuity and detail of fine vascular structures.

Notably, at the 2024 MICCAI conference, Dai et al. [139] introduced a 3D pad featuring multilayered, N-shaped, ultrasound-detectable lines for reconstructing 3D ultrasound volumes. This development demonstrates the transferability and adaptability of the optical pattern principle beyond PAI and into the ultrasound imaging domain, further validating its foundational potential across imaging modalities.

To conclude, this section has demonstrated that the optical pattern-based approach enables robust and accurate 3D photoacoustic volume reconstruction. Through the integration of novel hardware components and a modular, optimized software framework, the system is now sufficiently mature for clinical application, as explored in Part III. Beyond academic validation, commercialization strategies are actively being analyzed

within the scope of our current translational project funding (ForTra [202]).

Looking ahead, future research should aim to further mitigate remaining constraints of the current system. This includes the development of 3D pattern geometries to overcome orientation limitations, exploration of alternative or adaptive pattern designs, and the introduction of multicolor patterns. The latter would not only enhance the robustness and precision of pose estimation but also open the door to larger FOV stitching, thereby significantly extending the clinical and diagnostic potential of the method.

6 Multimodal image fusion

As shown in Part I (Fig. 3.2), PAI demonstrates strengths in spatial and temporal resolution as well as functional contrast. However, it inherently lacks sufficient penetration depth and anatomical contrast. Consequently, as an emerging imaging modality, PAI is not only positioned to be compared to established modalities such as MRI and CT, but also holds unique potential to complement them.

To realize this potential, it is essential to identify spatial relationships between modalities, thereby enabling image registration or fusion. Multimodal image fusion is generally achieved through one of the following strategies:

1) Integration within a single acquisition device;

2) Use of artificial landmarks or fiducials;

3) Identification of natural landmarks or image-derived features.

Approach (1) is typically applied to achieve PAI/US co-registration in both 2D and 3D, as the optical pattern coordinates used for photoacoustic acquisition are also utilized in the compounding of the 3D US volume. This section will primarily focus on approach (2), i.e., the use of artificial landmarks or fiducials. Since the optical pattern employed in PAI inherently acts as a set of spatial fiducials within the PAI domain, it is a natural extension to leverage this pattern for multimodal fusion with other imaging modalities.

To this end, additional fiducials were incorporated into the optical pattern, enabling spatial alignment across imaging systems by establishing a common coordinate framework. Rigid registration was then performed within this combined coordinate system. The following subsections describe the materials and methods used to implement this strategy, followed by quantitative validation in phantom experiments and qualitative demonstration in in vivo settings. Finally, the potential use of image-derived features for refined or non-rigid registration, outlining possible directions for future research are discussed.

6.1 Materials and methods

The core principle of pattern-based multimodal image fusion is as simple as it is elegant. By attaching artificial MRI- and/or CT-visible fiducials (referred to hereafter as fusion fiducials) to predefined locations on the optical pattern described in Sec. 5, a landmark-based registration becomes feasible (Fig. 6.1). These fusion fiducials (PinPoint®, Beekley

Medical, Bristol, CT, USA) are radiologically detectable and suited for small field-of-view applications. Their spatial relationship to the optical pattern establishes a fixed geometric linkage, enabling rigid registration between the fusion fiducials and the PAI pattern as elaborated in Sec. 2.2.2, particularly Eq. 2.22.

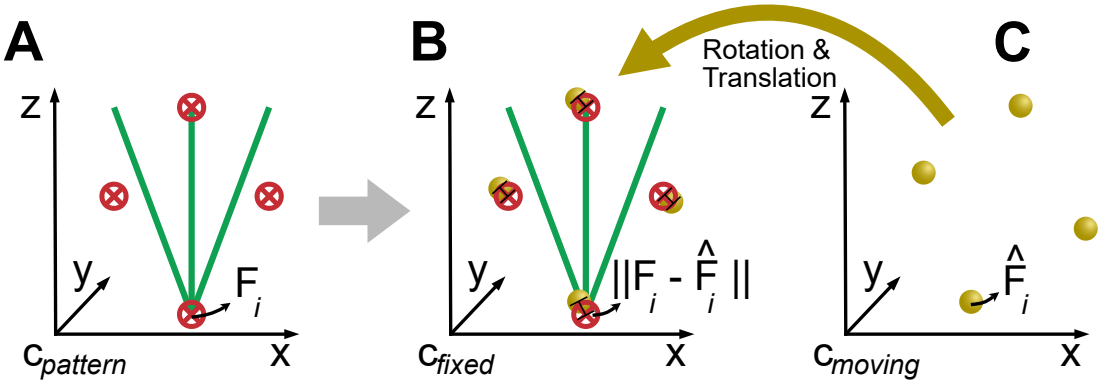


Figure 6.1: Fusion fiducials enable multimodal image fusion through rigid registration. **A** Additional artificial fiducial markers (red circles with crosshair) are placed at predefined positions F_i on the optical pattern within the pattern coordinate system $\mathbf{c}_{\text{pattern}}$. **C** These fiducials are visible to the second, non-photoacoustic imaging modality, which yields corresponding measured positions $\hat{F}_i$ (yellow spheres) in its own coordinate frame $\mathbf{c}_{\text{moving}}$. **B** A rigid transformation comprising rotation and translation is computed to align $\hat{F}_i$ with F_i, minimizing the Euclidean distance $\|F_i - \hat{F}_i\|$ across all fiducials (see Eq. 2.22, Sec. 2.2.2). The fixed image corresponds to the photoacoustic volume. Note that volume rescaling is not derived from the fiducials but performed independently during preprocessing using known voxel spacings of the respective modalities.

To ensure accurate and reproducible placement of the fusion fiducials, red circles were printed onto the pattern foil. These circles served as visual guidance markers and were experimentally verified to remain invisible in the resulting PAI image, ensuring no interference with optical signal acquisition. These pattern enhancements were developed in parallel with the hardware optimizations described in Sec. 5.1.1.1. For prototyping purposes, the original foil and ink as in the initial version of the system [185] were used. It is worth noting that multicolor screen printing is standard industrial practice, and the specific ink used for the guiding markers only needs to fulfill a minimal requirement, optical transparency above 650 nm. This condition is easily met by standard pigment formulations, thus enabling seamless integration into the newly developed system.

Two critical implementation parameters remain: the number and placement of fusion fiducials. From a theoretical standpoint, three linearly independent fiducials suffice to define and register two surfaces in three-dimensional space. This minimal configuration enables the computation of a rigid transformation between coordinate systems established by the fiducial geometry. To illustrate the geometric rationale, a single registered fiducial acts as an anchor point, but still allows arbitrary rotation of the two surfaces around this point. The addition of a second fiducial restricts rotational degrees of freedom, yet still permits mirroring of the surfaces, particularly in the left-right or up–down directions. The inclusion of a third fiducial resolves this ambiguity by fully constraining the transformation through the elimination of possible mirror symmetries.

Technically, in our current implementation, all three fiducials reside in a flat, coplanar geometry. This implies that up–down mirroring remains theoretically possible. However, this ambiguity is inherently resolved by the anatomy of the imaging context, where the orientation of tissue versus non-tissue (e.g., air or coupling medium) is unambiguous and fixed by definition.

Thus, while three linearly independent fiducials constitute the theoretical lower bound for rigid registration in three-dimensional space, this configuration should be viewed as the minimum viable arrangement. To more robustly determine the optimal number and placement of fiducials, we follow the four established principles of rigid body, point-based registration as proposed by West and Fitzpatrick [203]:

1) Avoid linear configurations of fiducials;

2) Minimize the distance between the centroid of the fiducials and the target:

3) Maximize the mutual distances between fiducials;

4) Increase the number of fiducials, bearing in mind that the benefit plateaus after approximately five or six fiducials.

These guidelines become particularly evident when analyzing the relationship between the expected squared target registration error (TRE^2) and the squared fiducial localization error (FLE^2) at a target point $\mathbf{r}$, as derived by Fitzpatrick et al. [204]:

$$\left\langle \mathrm{TRE}^2(\mathbf{r}) \right\rangle \approx \frac{\left\langle \mathrm{FLE}^2 \right\rangle}{N} \left(1 + \frac{1}{3} \sum_{k=1}^{3} \frac{d_k^2}{f_k^2} \right). \tag{6.1}$$

Here, N is the number of fiducials, d_k is the distance from the target $\mathbf{r}$ to the k^{th} principal axis of the fiducial configuration, and f_k is the root-mean-square (RMS) distance of the fiducials from that axis. The term $f_k^2/3$ represents the rotational contribution to the overall TRE. Linear arrangements violate principle (1), resulting in small f_k values and hence large TRE. Guideline (2) minimizes d_k, while (3) maximizes f_k. Finally, increasing the number of fiducials N as per (4) predictably reduces the TRE.

Following these design principles, we established the fiducial configuration illustrated in Fig. 6.1. Four fusion fiducials were symmetrically positioned around the optical pattern to minimize the distance between the fiducial centroid and the PAI volume (target) while maximizing their pairwise spacing. The decision to use four fiducials was guided by practical considerations namely, the limited surface area available at the imaging site and the physical footprint of the adhesive surface of the PinPoint® fiducials.

These fiducials are subsequently used to compute the rigid transformation between the PRF, defined by the optical pattern, and the CT/MRI coordinate system. Notably, volumetric rescaling is not derived from the fiducial configuration, but rather performed as a preprocessing step using the known voxel spacing of the respective input volumes. To quantitatively assess the feasibility and precision of the proposed registration strategy, a phantom experiment was conducted.

Phantom building To quantitatively assess the TRE of the proposed pattern-fiducial-based multimodal image fusion for PAI/MRI/CT, and to qualitatively evaluate the geometric consistency of the registration, a dedicated wire and point source phantom referred to as *Peter* (**P**hantom for **E**valuation of pat**TER**n-based fusion) was designed and constructed (Fig. 6.2).

As the base material for *Peter*, an agar-based tissue-mimicking phantom formulation was adopted following the protocol described in the supplementary material of Joseph et al. (2017), "Tissue mimicking phantom fabrication" [205], which is widely used in the PAI community. In brief, the fabrication consisted of four main steps: (1) Agar powder was dissolved by boiling in deionized water; (2) A scattering agent (Intralipid) and an absorber (Nigrosin) were each mixed with pre-warmed deionized water and added to the agar solution once it cooled to approximately 60°C; (3) The final solution was poured into a custom phantom box (inner dimensions: 14 cm × 18 cm); (4) The solution was allowed to cool and solidify.

Prior to pouring the agar solution into the phantom box, a flexible wire filled with a Nigrosin solution was suspended in the mold in two N-shaped configurations at depths of 1 cm (top layer) and 2.5 cm (bottom layer), with lateral extents of 3 cm and 1.5 cm,

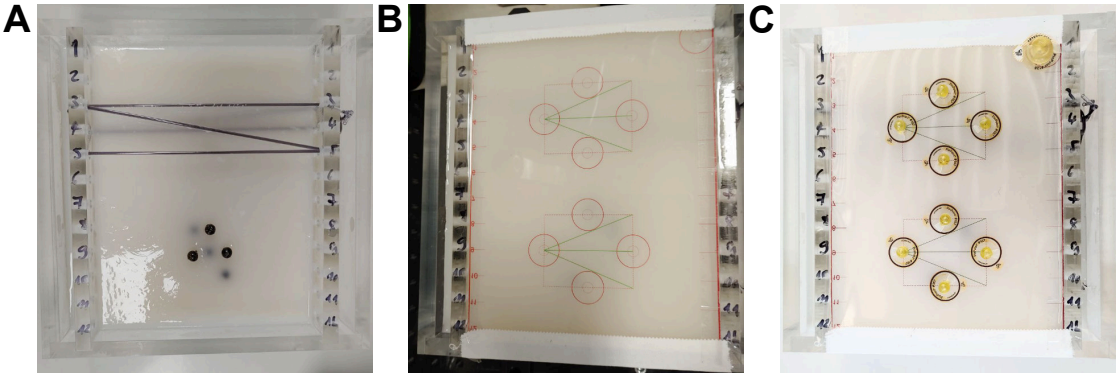


Figure 6.2: Construction steps of the custom validation phantom used for assessing the proposed optical pattern-based multimodal image fusion approach. **A** Interior view of the phantom during assembly, showing the N-shaped flexible wire structures filled with Nigrosin solution, the second layer of optically enhanced black spherical target markers, and the tissue-mimicking agar within the phantom container (inner dimensions: 14 cm × 18 cm). **B** Finalized phantom configuration prepared for photoacoustic imaging (PAI), featuring the applied fusion foil with embedded green optical pattern and red guiding fiducial print. **C** Setup for magnetic resonance imaging (MRI) and computed tomography (CT) imaging, where additional PinPoint® fiducials were adhered to the predefined locations on the foil to enable multimodal spatial registration.

respectively. The use of such embedded wire features aimed to provide high-contrast targets in PAI, although it was initially uncertain whether sufficient contrast would be observed in MRI or CT.

Once the agar solution was poured to the level of the lower wire structure and allowed to partially cool, the first layer of point targets was embedded at the opposite end of the phantom. As point targets, commercial spherical PinPoint® markers (6 mm inner diameter) were used. Preliminary experiments indicated that these markers did not produce sufficient photoacoustic signal. Therefore, they were coated with black permanent marker ink (edding®) to enhance their optical absorption and improve PAI visibility.

This process was repeated for a total of nine spherical targets, arranged in three layers. Each layer was successively embedded and covered with additional agar solution. After cooling and solidification, the fusion foil was placed on top of the phantom to facilitate PAI measurements. For MRI and CT acquisitions, additional PinPoint® markers were adhered to the designated positions on the fusion foil, guided by pre-printed circular markers.

In vivo setup To provide an in vivo proof of concept, a calf measurement was performed using both an MSOT system and a 3T MRI scanner (Siemens) at the National Center for Tumor Diseases (NCT). In order to avoid unnecessary radiation exposure, only PAI–MRI fusion was performed. However, based on the quantitative results from the phantom validation in Sec. 6.2, comparable registration accuracy can be expected for PAI–CT fusion.

For this study, the optical pattern was secured to the author's calf using two to three layers of medical adhesive tape to ensure stability during transportation. This precaution was necessary as the distance between the MSOT and MRI imaging facilities spans several hundred meters. The imaging protocol was identical to that employed in the phantom experiments: the MSOT system was used in combination with the pattern-specific acquisition and reconstruction workflow (Sec. 5.1.3.2), and ultrasound gel was applied on both sides of the fusion foil to maintain acoustic coupling.

6.2 Validation and results

This section presents the validation results and qualitative in vivo imaging outcomes. We begin with the quantitative assessment of the fiducial-based registration, visual inspection of multimodal alignment, and target registration error (TRE) analysis using the *Peter* phantom. This is followed by the demonstration of in vivo multimodal fusion, concluding with an outlook on potential non-rigid, image-based registration refinements.

Phantom Following the fabrication process, the *Peter* phantom was imaged from both the point target side and the N-wire target side. The imaging was performed using the MSOT system in accordance with the acquisition and processing protocol tailored for the optical pattern (Sec. 5.1.3.2). Ultrasound gel was applied both below and above the fusion foil to ensure adequate acoustic coupling. The foil was additionally stabilized using adhesive medical plaster to minimize motion artifacts during PA data acquisition.

Subsequent to the PAI scans, the fusion foil was carefully cleaned and dried, and the fusion fiducials were adhered to the predefined positions indicated on the foil. This allowed for precise spatial correlation with subsequent imaging modalities. CT measurements were performed using an in-house Siemens scanner with 0.6 mm isotropic voxel spacing. MRI was conducted on an in-house 3T Siemens system, utilizing an imaging protocol yielding 0.9 mm isotropic resolution.

For landmark-based registration, both the fiducial markers and internal target spheres were manually segmented. A threshold-based segmentation approach was employed for each of the 3D volumes (PAI, CT, and MRI), where modality-specific intensity thresholds were defined to isolate the fiducials and target structures. These threshold parameters were optimized manually for each modality to ensure reliable and consistent landmark extraction.

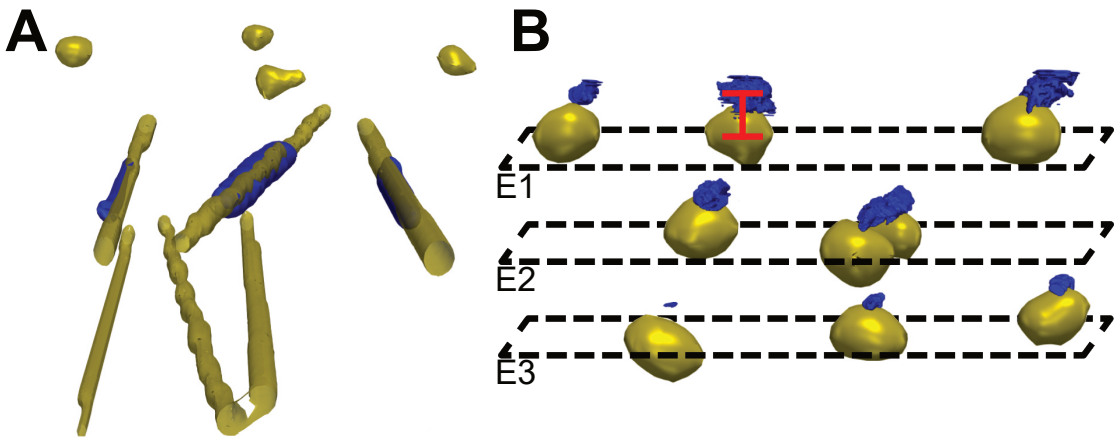


Figure 6.3: Three-dimensional rendering of the *Peter* phantom, combining computed tomography (CT) data (yellow) and pattern-based photoacoustic imaging (PAI) data (blue). **A** Visualization of the fusion fiducial markers affixed to the phantom surface and the two embedded N-shaped wire layers. Note that no PAI signal was detectable at the depth of the second (deeper) N-wire structure due to the limited penetration depth of the modality. **B** Close-up view of the embedded spherical point targets positioned at three distinct depth levels (E1–E3). The red schematic overlay illustrates the inherent spatial offset between the designated PAI target (black-painted surface of the marker), and the actual registration target defined as the center of mass of the spherical marker. Importantly, the spherical geometry visible in the CT volume does not correspond to the marker fluid but rather to an internal air inclusion. The lack of contrast in Hounsfield units (HU) between the marker fluid and the surrounding agar rendered the fluid indistinct in the CT image (Appendix Fig. A9).

Since the fusion fiducials were designed and commercialized as MRI-visible markers, no measurable contrast was observed between the fiducial marker fluid and the surrounding agar in the CT data (Appendix Fig. A9). Nevertheless, both the fusion fiducials and internal air inclusions within the spherical target markers, as well as the embedded N-shaped wires containing Nigrosin solution, were clearly distinguishable

in the CT volumes (Fig. 6.3). Nonetheless, the primary focus of the following evaluation is on MRI–PAI image fusion, although similar observations were made for the CT–PAI case.

Following threshold-based segmentation of the fiducials, the transformation between the fixed (PAI) and moving (MRI) volumes was computed. Landmark registration was based on the centroids of the segmented fiducials, and the transformation was initialized using the `sitk.LandmarkBasedTransformInitializer` function. The resulting rigid registration is illustrated in Fig. 6.4.

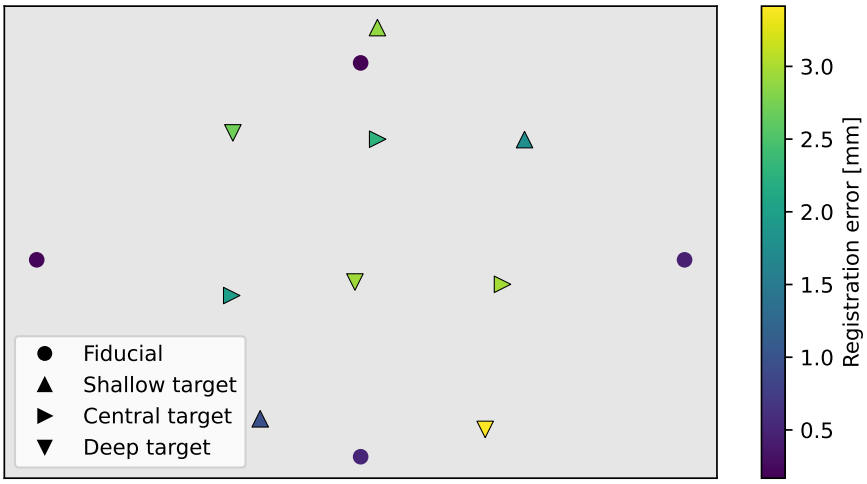


Figure 6.4: Post-registration x-z projection of all nine target markers (triangles) and four fiducial markers (●) for magnetic resonance imaging (MRI) photoacoustic imaging (PAI) image fusion. Triangle orientation encodes assigned depth levels: shallow (↑), central (→), and deep (↓), corresponding to E1-E3 depth categories (Fig. 6.3). Marker color indicates the local registration error.

Quantitative evaluation of the fiducial registration yielded a FRE of 0.37 mm (Eq. 2.22), with a minimum and maximum error (Frobenius norm [206] from `numpy.linalg.norm`) of 0.17 mm and 0.52 mm, respectively (CT: 0.82 mm [0.32-1.28 mm]). To compute the TRE, an additional depth offset was applied. This correction accounts for the discrepancy in target definition across modalities: in the PAI image, only the painted top surface of the spherical marker (radius = 3 mm) is visible, whereas in the MRI image, the defined target corresponds to the geometric center of the marker (Fig. 6.3 **B**). Incorporating this correction, the resulting TRE was 2.54 mm (Fig. 6.4), with values ranging from 0.96 mm to 3.41 mm (CT: 3.33 mm [0.56-5.33 mm] Appendix Fig. A10). A depth-dependent trend was observed, whereby targets located deeper within the phan-

tom exhibited higher TRE values, consistent with the expected increase in registration uncertainty due to greater spatial separation from the fiducial plane.

In vivo To qualitatively demonstrate the feasibility and limitations of the proposed pattern-based multimodal image fusion, an in vivo case study was conducted, starting with a 3D PAI measurement.

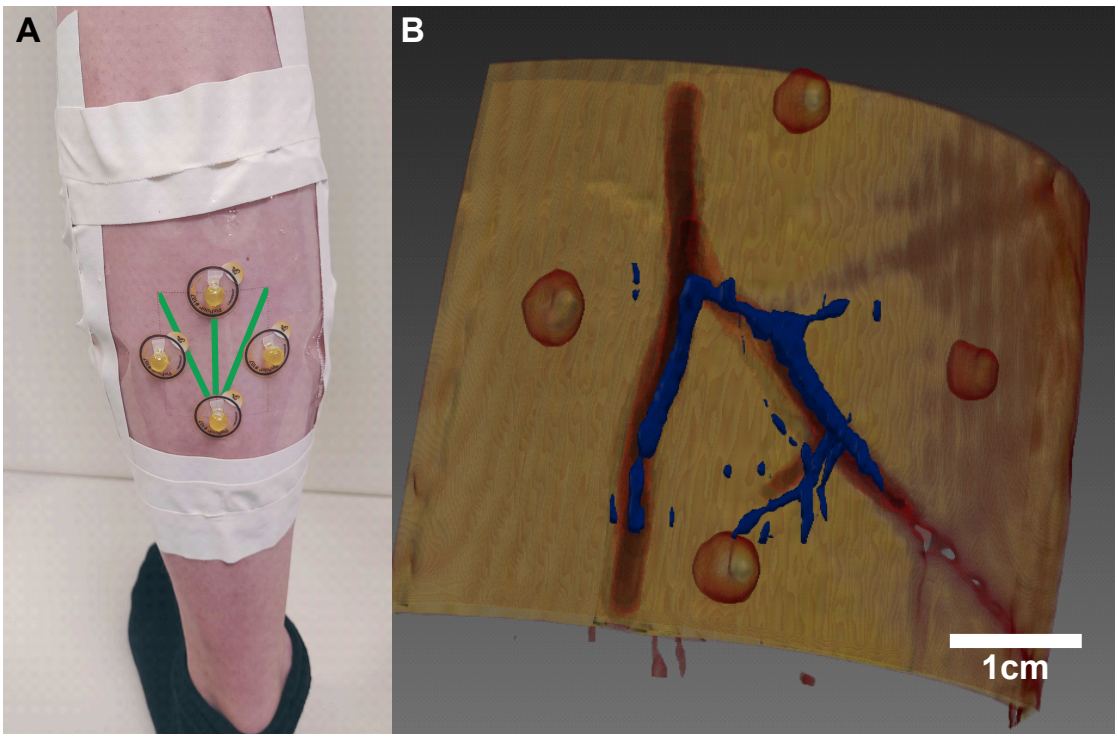


Figure 6.5: **A** In vivo setup illustrating the application of the optical pattern (graphically highlighted in green), fiducial markers (yellow spheres), and fusion foil on the posterior calf surface. To minimize motion-induced misalignment during the approximately 300 m transit between the photoacoustic imaging (PAI) and magnetic resonance imaging (MRI) facilities, the pattern was secured with multiple layers of medical tape. **B** Three-dimensional rendering of the fused PAI–MRI dataset acquired from the same subject. The MRI image is shown in yellow/red, while the overlaid PAI volume is depicted in blue. The PAI data were thresholded and cropped to exclude skin layers and highlight superficial vasculature through vessel segmentation. The fused image reveals a well-aligned superficial vessel, demonstrating the complementarity of both modalities: MRI provides detailed anatomical context, whereas PAI delivers high-resolution functional detail, particularly of vascular structures.

Upon arrival at the MRI suite (NCT facility), the fusion fiducials were carefully positioned at their predefined locations on the fusion foil. MRI was then performed with an isotropic resolution of 0.9375 mm using a T1-weighted Dixon sequence. The subject was positioned face-down such that the posterior side of the calf where the pattern and fiducials were attached faced upwards (setup illustrated in Fig. 6.5 **A**). A fused 3D rendering of the MRI and PAI volumes obtained from this in vivo measurement is displayed in Fig. 6.5 **B**.

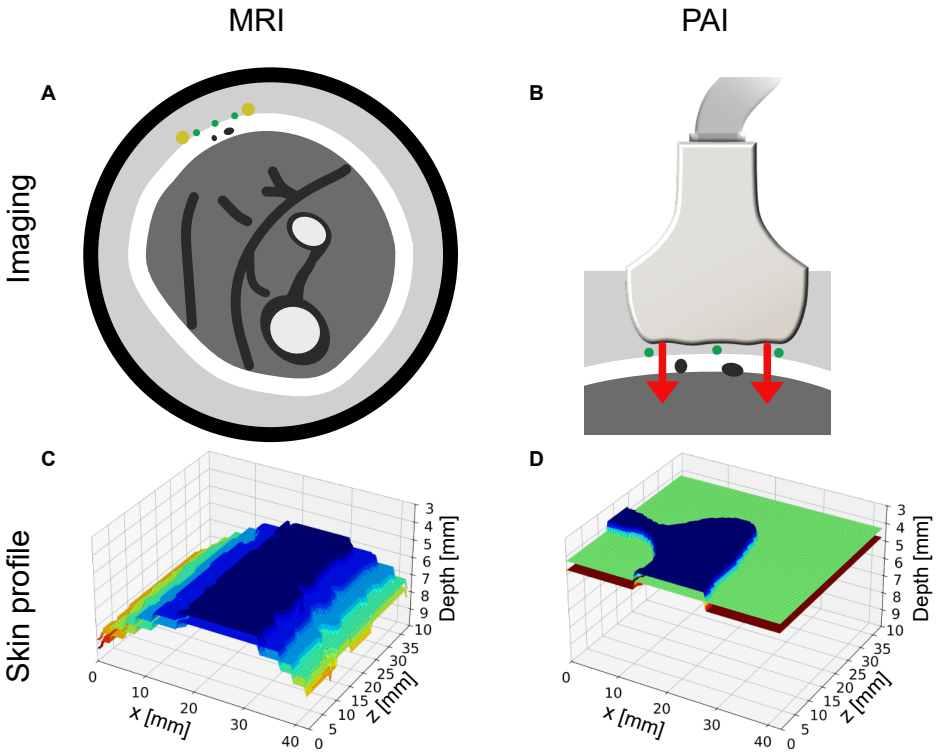


Figure 6.6: **A** Iconized 3D rendering of the subject's calf acquired during magnetic resonance imaging (MRI). Superficial vessels within the subcutaneous white fat layer are shown in gray, while fusion fiducials and the optical pattern are visualized in yellow and green, respectively. **B** Photoacoustic (PA) field of view (FOV) of the same calf region, displaying the optical pattern (green). Red arrows indicate probe-induced mechanical compression, deforming the tissue towards a flat surface during PA acquisition. PA probe, illustrated by Tom Rix. **C** Threshold-based profile of skin surface curvature extracted from the MRI volume, illustrating the natural rounded contour of the calf. **D** Corresponding skin profile derived from the PAI volume, appearing predominantly flat due to compression, besides minor thresholding artifacts. A direct comparison of unprocessed 2D image slices depicting the deformation-induced shape differences is provided in Fig. 6.7.

However, a key limitation of the presented pattern-based multimodal image fusion

approach was identified (Fig. 6.6). During the PAI acquisition, the applied probe pressure flattened the skin surface (Fig. 6.6 **B**/**D**). In contrast, the MRI scan acquired with the subject lying relaxed depicted the calf in its natural, slightly curved anatomical shape (Fig. 6.6 **A**/**C**). This discrepancy is reflected in one of the foundational assumptions of the pattern-based 3D reconstruction method: a flat imaging surface. As a result, non-rigid skin deformation introduced during PAI acquisition led to spatial misalignment between PAI and MRI volumes, particularly near the skin boundary.

To address this misalignment, a non-rigid correction (`sitk.BSplineTransform` [195]) was applied to match the skin segmentation masks after an initial rigid registration. This approach yielded visibly improved co-registration but also introduced potentially physiologically meaningless deformations of vascular structures (Fig. 6.7). These observations highlight both the importance of deformation-aware fusion strategies and the potential need to incorporate non-rigid registration techniques in future applications.

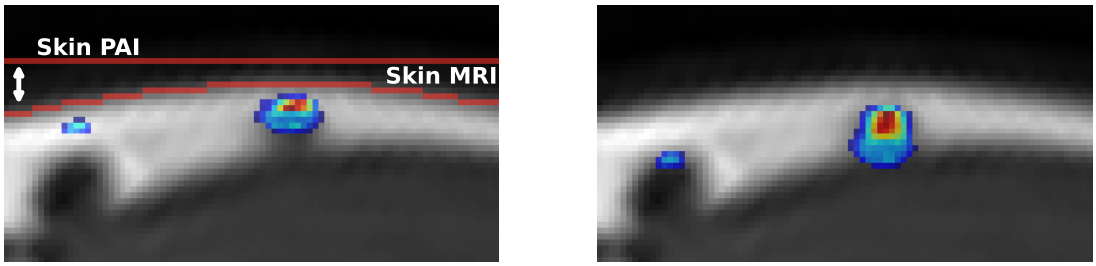


Figure 6.7: Magnetic resonance imaging (MRI) image (grayscale) overlaid with photoacoustic (PA) data (color; transparency threshold set to 15 arb. units), registered using the pattern-based fusion approach. **A** Pre-correction fusion result, showing a mismatch in skin curvature caused by deformation during PA acquisition. The red contours correspond to skin surface profiles extracted from the MRI and photoacoustic imaging (PAI) volumes (as shown in Fig. 6.6 **C**/**D**). **B** Post-correction result after image-based non-rigid registration refinement, illustrating improved alignment of vascular structures.

6.3 Discussion

This chapter presented an extension to the optical pattern-based approach for 3D PAI, enabling multimodal image fusion through landmark-based rigid-body registration. This was achieved by integrating additional fiducial markers at predefined locations on the optical pattern. The proposed method allows spatial co-registration of PAI with established imaging modalities, such as MRI and CT.

Validation of the method was carried out using a custom-designed phantom incorporating both N-shaped wire targets and spherical point targets. Qualitative assessment confirmed geometric consistency, and quantitative analysis yielded a mean TRE of less than 3 mm. This level of accuracy aligns with values reported as acceptable for multimodal image registration in the literature [207]. Furthermore, the concept was successfully demonstrated in vivo through the fusion of PAI and MRI images of the author's calf.

However, this in vivo validation also revealed a key limitation of the proposed method. The pressure exerted during PAI acquisition leads to the deformation of superficial tissues, causing a flattening of the skin surface. In contrast, the MRI acquisition performed without mechanical compression preserves the natural curvature of the tissue. Since the current pattern-based 3D reconstruction assumes a flat imaging surface, this non-rigid deformation introduces spatial misalignment. To address this, an image-based registration refinement was explored. While rigid correction by adjusting depth alignment improved fusion quality, full non-rigid deformation modeling (e.g., based on ultrasound elastography) was beyond the scope of this work.

Nonetheless, the relevance of non-rigid registration in this context remains debatable. Ferrante et al. [52] argue that rigid approximations are "expressive enough" for clinical use, being both more interpretable and robust, while "non-rigid and elastic models might lead to solutions which are correct from a geometrical point of view but they are not anatomically meaningful" [52]. This perspective underscores another limitation of the current study: the absence of a defined clinical use case. Without a task-specific benchmark, it remains unclear whether the achieved precision and accuracy are sufficient for any particular application.

One initial attempt at clinical translation involved fusing PAI with X-ray images of porcine models exhibiting intestinal obstruction due to cystic fibrosis (CF). However, these efforts were unsuccessful due to equipment malfunction and the presence of severe air artifacts in the gastrointestinal tract, which impeded evaluation.

Another practical constraint lies in the accuracy of fiducial placement. Misalignment of the fusion fiducials directly increases the FLE, which in turn limits the achievable TRE as per Eq. 6.1. In practice, deviations of several millimeters seem plausible during phantom construction, partially due to the design of the commercial PinPoint® fiducials. The actual marker (yellow sphere) is not centered within the adhesive surface encircled by a black ring (Fig. 6.2 **C**), a design acceptable for clinical localization tasks but suboptimal for high-precision registration.

To the best of my knowledge, the only comparable work in the literature is that of

Park et al. [161], who developed a marker-based MRI–PAI registration method using optical tracking. Their system achieved a TRE of approximately 4 mm. While optical tracking systems support larger and more flexible FOVs, they require calibration and line-of-sight stability and are more susceptible to patient motion. In contrast, the optical pattern approach presented here is calibration-free and robust against such limitations.

Beyond this, most PAI registration research has focused on small animal imaging, where non-rigid deformations are less problematic due to the use of water-bath-based scanner designs with no direct mechanical contact [155]. Nonetheless, Manohar et al. [162] identified multimodal integration as one of eight key challenges for the advancement of "super phantoms," underscoring its importance in the broader context of medical imaging research.

In conclusion, this work demonstrates that the optical pattern-based approach for 3D PAI can be extended to enable rigid multimodal image fusion, with a mean TRE below 3 mm in phantom experiments. While no clinical application was successfully validated, the achieved precision is promising and suggests future clinical utility. Identifying and validating such a use case should be a primary focus of subsequent investigations. Based on the intended application, further refinement via image-based registration either rigid, as shown in the temporal registration context (Sec. 5.1.2.4), or non-rigid may be warranted. Multimodal fusion also offers solutions to broader challenges in the PAI field: from enabling comparative benchmarking and resolving the lack of anatomical landmarks in PAI-only imaging, to improving clinical interpretability, especially in deeper tissue regions where PAI alone is limited.

7 Digital twin-based failure analysis

PAI enables functional imaging by exploiting the optical absorption properties of tissue chromophores. However, one of the fundamental challenges in PAI is that the measured signal does not directly represent optical absorption or chromophore concentrations. Instead, it is influenced by a cascade of physical processes including fluence distribution, noise and acoustic effects such as limited view, making it difficult to isolate true biological effects from imaging-induced distortions. This "lack of ground truth for the provided measurements, hinder[s] the accurate assessment of biological effects, hence leveraging biological context. However, the hybrid US-PAI technology provides us with reliable information about tissue morphology. This morphological data, when coupled with literature-derived optical and acoustic properties, enables us to develop digital tissue twins, potentially circumventing the ground truth limitation. With this tissue twin model, we can simulate different biological effects with exact knowledge of the ground truth, e.g. oxygenation. Leveraging the ground truth knowledge, we can analyze how the biological change influences the photoacoustic device twin measurement." [93]

This section outlines the modeling methods used to construct a digital twin of the PAI pipeline and presents a failure analysis of a clinical study based on simulated 2D PAI data, including a qualitative discussion of the framework's current limitations. The clinical motivation, as well as the experimental details and implications of this work, are discussed in Part III (Sec. 8). Here, we restrict the discussion to digital twin modeling and its role in probing potential failure modes of oxygenation quantification.

To contextualize the modeling effort, a brief summary of the clinical application is provided. The study investigated malignant lymph nodes in patients with head and neck cancer undergoing fractionated radiotherapy. Longitudinal 2D PAI measurements were obtained prior to treatment, after the 9^{th} fraction, and three months post-therapy. Surprisingly, the oxygenation levels estimated from the PAI data declined throughout and after the treatment period. An observation that contradicts both domain expectations and clinical consensus [208]. These findings prompted a re-evaluation of the photoacoustic quantification pipeline, with a particular focus on its failure characteristics under realistic but uncontrolled conditions using digital twin-based modeling.

The methods and core results reported here are based in large part on my first-author publication [93], which introduced the digital twin-based failure analysis in the context of therapy response monitoring.

7.1 Materials and methods

This section outlines the methodological framework underlying the development of the digital tissue and device twin modeling in PAI, illustrated through the generation of a data set used for the digital twin-based failure analysis.

As previously described in Part I (Sec. 2.1.4), the complete synthesis pipeline comprised three primary stages: (1) volume generation, (2) property assignment, and (3) optical and acoustic modeling. The central innovation of the presented digital tissue twin lies in its capacity to incorporate patient-specific characteristics, such as anatomical geometry and pathological conditions, into the modeling process (Fig. 7.1). This approach enabled the investigation of the intrinsic limitations of PAI image formation in the presence of biologically relevant effects simulated by the model.

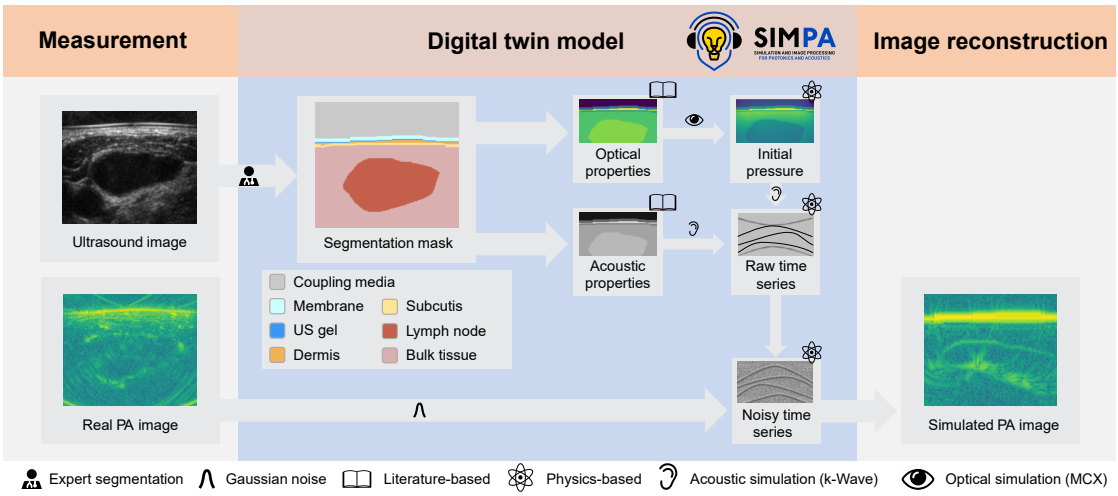


Figure 7.1: "A comprehensive simulation pipeline enables the generation of tissue digital twins. To obtain synthesized images that are geometrically consistent with our measured data, ultrasound images (US) are first converted into a semantic representation (segmentation mask) by clinical experts. The segmentation masks are then annotated with plausible optical and acoustic tissue properties obtained from the literature. Finally, existing methods for photon (Monte Carlo eXtreme - MCX) and acoustic wave (k-Wave) propagation as well as realistic noise generation, provided by the SIMPA toolkit [35], are applied to generate a digital twin photoacoustic (PA) image." [93] Figure taken from [93] under CC BY 4.0 [120].

To enable the generation of patient-specific tissue twins, "expert physicians converted the measured data [of all three timepoint (Sec. 8)] into semantic segmentation masks [using MITK [209]], resulting in a labeled dataset with four tissue classes (dermis,

subcutis, bulk tissue, and lymph node) and three material classes (coupling media, membrane, and ultrasound gel), as illustrated in Figure 7.1." [93] Post-processing steps included the merging of related classes, the closing of segmentation gaps, and the spatial extrapolation of 2D segmentation masks into the lateral dimension to construct a full 3D tissue model.

To generate realistic tissue twins for simulation, parameter sampling was carefully designed to reflect physiologically plausible variability across and within the tissue classes. For each simulated volume, tissue class-dependent optical (μ_a, μ_s, g) and acoustic (density, speed of sound, acoustic attenuation) properties were sampled from literature-based distributions. These distributions were based on the molecular composition of the respective tissues as compiled within the SIMPA toolkit, "based, among others, on work by Jacques et al. (2013) [210] ([Appendix Tabs. B1, B2, B3). To mimic disease characteristics [in the lymph nodes], different [blood volume fraction] blood volume fraction (BVF) and sO_2 values were assigned to every instance of the SIMPA lymph node object [(Fig. 7.2)], specifically simulating three disease states: malignant (BVF 4 $\pm$ 2%, sO_2 43 $\pm$ 15%), benign (BVF 14 $\pm$ 7%, sO_2 73 $\pm$ 15%), and an artificial intermediate state (BVF 9 $\pm$ 4.5%, sO_2 58 $\pm$ 15%) [(Fig. 7.2)], based on single reflectance spectroscopy measurements by Bugter et al. (2021) [211]" [93]. Importantly, disease-state-dependent parameters (e.g., blood volume fraction and sO_2) were only varied in the lymph node class and not in the surrounding tissue.

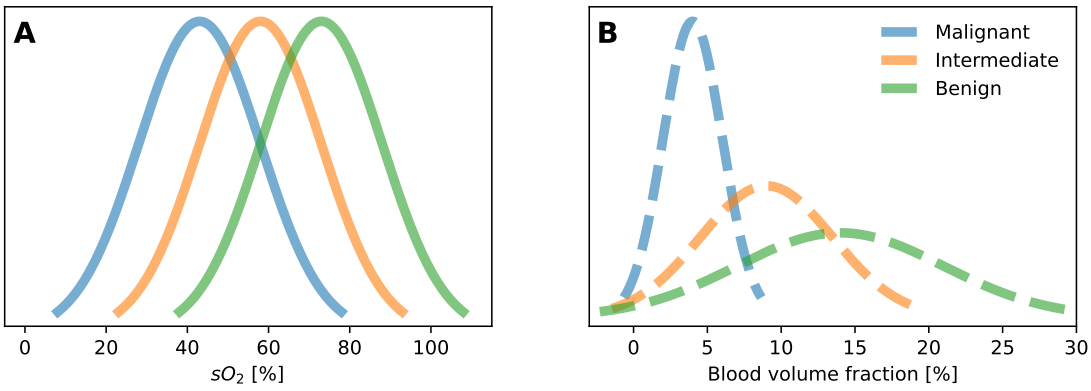


Figure 7.2: Distribution assumptions of oxygen saturation (sO_2) and blood volume fraction (BVF) for the three modeled disease states: malignant, benign, and an artificial intermediate condition. (A) shows sO_2 distributions, while (B) illustrates the corresponding BVF distributions. Parameter values for malignant and benign lymph nodes were derived from single reflectance spectroscopy measurements [211], and the intermediate state was interpolated accordingly.

"After volume generation, optical and acoustical simulation additive Gaussian noise was applied to the time series data using the SIMPA GaussianNoise adapter, to imitate a realistic noise level comparable to the real data." [93] The GaussianNoise module superimposed zero-mean Gaussian-distributed noise onto each simulated time series (i.e., the recorded acoustic pressure signals per voxel), with the standard deviation empirically set to replicate the signal-to-noise ratio observed in in vivo measurements. The noise model was applied globally to all tissue regions, assuming a spatially homogeneous background noise floor as expected from system electronics and thermal noise in typical handheld imaging setups.

7.2 Validation and results

This section presents the experimental results of the digital twin-based failure analysis. All experiments were conducted using the previously described dataset, which monitored lymph node oxygenation in the head and neck region during radiotherapy across three distinct time points, representing different disease states.

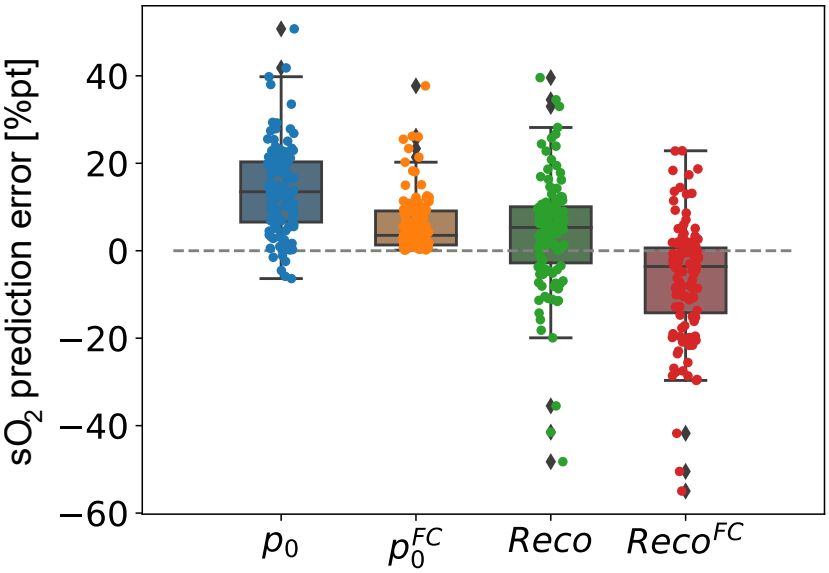


Figure 7.3: A data point represents the mean oxygen saturation (sO_2) prediction error for both the initial pressure distribution (p_0) and the reconstructed photoacoustic (PA) images (Reco), with and without fluence correction (FC), evaluated per lymph node. Results are shown for all simulated lymph nodes. Applying fluence correction at the level of p_0 reduces the typical estimation error of sO_2, resulting in more accurate estimates. In contrast, applying the same correction after acoustic simulation and image reconstruction ($Reco^{FC}$) leads to overcorrection and a systematic underestimation of sO_2 values.

Failure analysis In the context of analyzing the theoretical sources of error and limitations in sO_2 estimation via PAI, wavelength-dependent fluence variation is typically regarded as a major source of uncertainty [40].

However, for the specific experimental design and anatomical configuration addressed in this study, the digital tissue twin provided additional insights. As shown in Fig. 7.3, applying fluence correction (division by the simulated fluence map) directly to the initial pressure images, i.e., after the optical simulation step but before acoustic propagation, resulted in improved estimation performance, as reflected in lower sO_2 estimation errors (deviation from the ground truth sO_2).

In contrast, applying the fluence correction after the complete simulation pipeline, including both optical and acoustic modeling as well as image reconstruction, led to sO_2 estimation errors of similar magnitude, but with reversed sign.

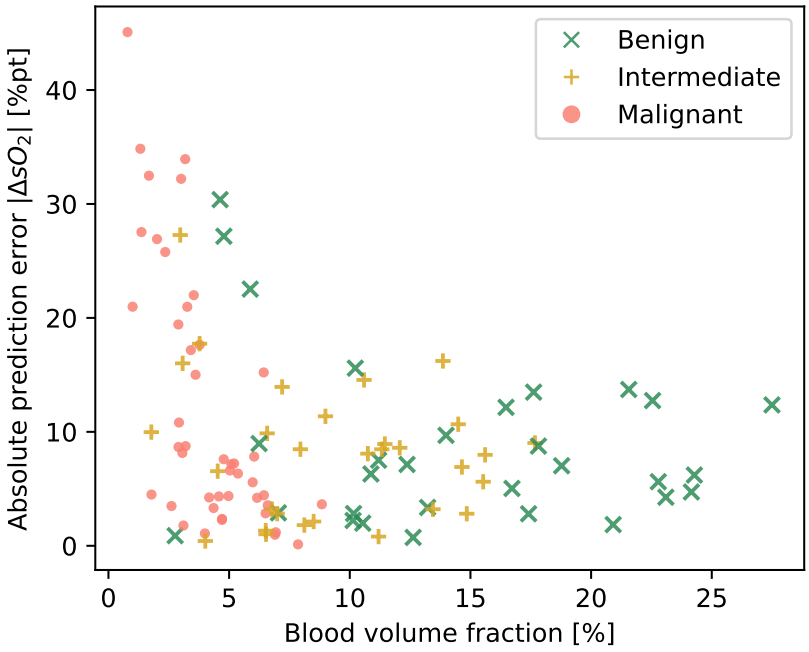


Figure 7.4: "Digital twin analysis reveals a strong link between oxygen saturation (sO_2) prediction error (y-axis) and low blood volume fraction (BVF) (x-axis) for benign, intermediate, and malignant lymph nodes. Every [data point] represents the mean absolute sO_2 error of one digital twin scan with the sO_2 ground truth defined by the assigned optical input properties." [93] Figure taken from [93] under CC BY 4.0 [120].

These results indicate that, beyond wavelength-dependent fluence effects, additional

physiological and physical factors contribute significantly to the sO_2 estimation error observed in this study. Here, "the digital twin analysis revealed that low blood volume fraction (BVF), as encountered in malignant nodes, can lead to particularly high sO_2 prediction errors (Fig. 7.4). In other words, the measured sO_2 values cannot be trusted in regions of low BVF" [93].

7.3 Discussion

"A digital twin model shed light on PAI-specific challenges in signal generation from malignant lymph nodes, particularly revealing low BVF as a factor associated with high prediction errors in sO_2" [93].

However, the modeling remains far from perfect. Each processing step introduces new assumptions, some of which are incorrect, leading to accumulated inaccuracies throughout the simulation pipeline. This ultimately contributes to the previously reported domain gap [212], which hinders a one-to-one translation of digital twin assumptions and results, into clinical interpretation. These inaccuracies begin with the volume generation step, which typically involves large, homogeneous tissue regions, whereas real biological tissue exhibits significant heterogeneity. Furthermore, it remains unclear how accurate the assigned optical and acoustic properties are, due to the lack of labeled reference data, a challenge also addressed with a learning-to-simulate approach [213].

Another limitation lies in the fundamental restriction that phenomena cannot be modeled if they are not known. For example, the exact spatial and temporal molecular changes occurring across diverse tissue types in the head and neck region, especially under therapeutic conditions such as radiotherapy.

Importantly, the study found that fluence correction applied at the initial pressure level led to improved sO_2 estimation, whereas applying the same correction post-reconstruction failed to yield meaningful improvements due to error accumulation introduced by both modeling and reconstruction processes. This indicates that, under the given morphological and signal-processing conditions, post hoc fluence correction at the image level may not effectively compensate for most relevant artifacts and can even introduce new biases. However, even in post-reconstruction cases, the fluence-corrected spectra appeared more characteristic, despite only minor effects on sO_2 accuracy.

Still, the most significant limitations found in the current study are the inability to model dynamic changes in size and depth of the nodes, the presence of unmodeled background signal, and the use of a simplistic noise model. Additionally, "malignant

lymph nodes pose an aggravating factor due to their low BVF and high acoustic attenuation, resulting in particularly weak measurable photoacoustic signals. The importance of this effect was revealed through the presented digital twin model." [93].

To overcome barriers in PAI towards clinical translation [6], a thorough understanding of system limitations as provided by digital twin-based failure analysis is crucial. However, the current approach lacks realism, and there is a pressing need for more accurate simulation techniques or improved domain transfer methodologies [212].

To conclude, the digital twin tool is helpful in identifying potential causes of failure during study planning and analysis. It provides guidance on where PAI might work effectively and helps delineate its inherent limitations.

- Part III: Clinical applications -

Concept overview

The preceding part introduced the core technical contributions, particularly the development of a robust framework for optical pattern-based 3D PAI and a novel digital-twin-based failure analysis methodology. These foundational components are instrumental in enabling the translational efforts described in the following chapter, which focus on the clinical application of PAI. This part of the thesis (Part III) explores the application of PAI in two distinct clinical use cases. Beyond demonstrating the feasibility of the proposed applications, each study also highlights inherent limitations and open challenges in the field ranging from the complexity of capturing dynamic biological processes with PAI, to the definition of clinically meaningful ROIs.

The two investigations covered are:

i) Longitudinal monitoring of radiotherapy response in malignant lymph nodes of patients with head and neck (HN) cancer (Sec. 8), including the use of contextual metrics such as the tumor-to-muscle ratio to evaluate tissue oxygenation in relation to surrounding anatomical structures;

ii) Detecting peripheral artery disease (PAD) using 3D PAI, with an emphasis on the diagnostic value of spatial context in identifying perfusion-related abnormalities (Sec. 9).

Each study is presented in a consistent structure, beginning with the design and methodology, followed by implementation, results, and discussion.

8 Monitoring of tissue response to radiotherapy with photoacoustic imaging

This chapter is primarily based on my first-author publication: "Photoacoustic imaging for monitoring radiotherapy treatment response in head and neck tumors," published in *Scientific Reports* (2025) [93]. The study has already been briefly introduced in the context of the digital twin-based failure analysis (Part II Sec. 7).

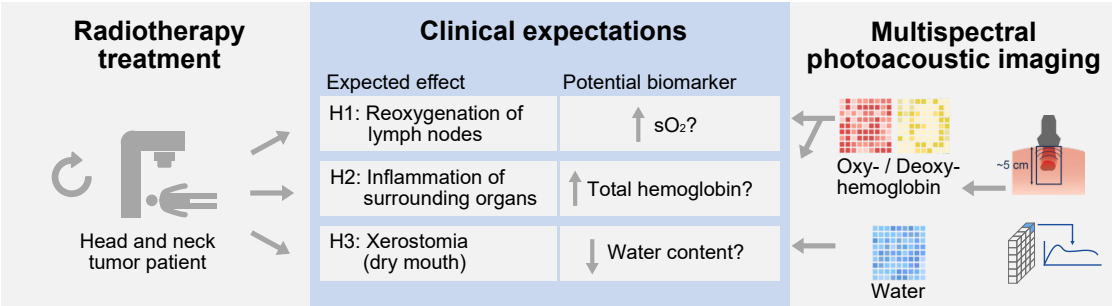


Figure 8.1: "Mission of this work: to investigate whether photoacoustic imaging can measure changes in tissue caused by radiotherapeutic treatment. The specific focus was to analyze whether clinically expected short- (H1) and long-term (H2 & H3) effects resulting from radiotherapy manifest in photoacoustic measurements." [93] Figure taken from [93] under CC BY 4.0 [120].

"Given the gap in the literature [(missing early response measurement of RT treatment: Part I Sec. 3.3.2)], this study aims to explore the potential of PAI in assessing the response to RT in malignant lymph nodes of head and neck tumors in vivo. To achieve this objective, we conducted the first explorative patient study on head and neck tumor patients who underwent repeated imaging sessions throughout RT treatment with PAI. Our approach involved matching the expected treatment effects with potential biomarkers extractable using PAI (Fig. 8.1). Specifically, we hypothesized the following effects of RT:

> (H1) An early molecular increase in oxygenation in lymph nodes resulting from treatment.
>
> (H2) A late increase in total hemoglobin, indicative of inflammation.
>
> (H3) A late decrease in saliva production and, consequently, reduced water content in salivary glands." [93]

In addition to the published results, this chapter includes novel and extended analyses beyond those presented in the paper. These include:

- A more detailed evaluation of the LMM results, providing further insight into variability across patients and lymph node regions.
- An extended analysis of the correlation between blood sample derived total hemoglobin (tHb) and the MSOT signal.
- A brief qualitative assessment of organ-specific “fingerprints” in multispectral PA imaging, highlighting their potential for semantic segmentation.

8.1 Materials and methods

The study was conducted using a non-CE-certified version of the MSOT system (Part II Sec. 5.1.3.2), as it took place prior to device certification and under the regulatory framework of the former medical device directive.

8.1.1 Ethics

"This study adheres to the Helsinki Declaration guidelines, was approved by the ethics committee of the Medical Faculty of Heidelberg University and was registered at clinicaltrials.gov (ID NCT04437030). All procedures were conducted with full informed consent from the participants." [93]

8.1.2 Study design, inclusion criteria, and participants

"The PAI examinations were conducted on patients diagnosed with squamous cell carcinoma at various locations in the HN region, including the floor of the mouth, tonsil, and base of the tongue who received definitive radiotherapy without prior surgery. The patients had easily accessible lymph node metastases, primarily in levels I-III according to Gregoire’s classification [214] (further tumor-related details see Tables B4 and B5 in the [Appendix]). Originally, the study included 15 young healthy volunteers to train the physicians and test the measurement protocol, alongside 15 patients. Within the patient cohort, there were dropouts due to health conditions. Two patients were direct dropouts from the study, one patient died after radiotherapy, and one patient could no longer be examined at the third appointment due to declining health. Finally, the analysis included 11 complete patient datasets (three time points) and two incomplete patient datasets (two time points). 11 of the patients received

moderately hypofractionated regimen (32 x 2,2 Gy with 70,4 Gy on tumor and 57,6 Gy on lymph node area) - one patient hypofractionated (due to health state) with 17 x 3 Gy (51 Gy), and one patient combined treatment with protons (57.6 Gy relative biological effectiveness (RBE) in weekly 5 x 1.8 Gy RBE single doses with simultaneous integrated boost [...] to the lymph nodes and the base of the tongue up to 64 Gy RBE and [...] to the tumor up to 70.4 Gy RBE). Also, 11 patients received concomitant systemic therapy (10x Cisplatin, 1x Cetuximab 5-6 Cycles). All patients underwent follow-up examination at six to eight weeks after the end of therapy by CT or MRI, which allowed us to assess tumor response in another modality. Lymph node targets were CT- or MRI-graphically measured in two planes before radiotherapy and after the period described above. As a reference before radiotherapy, we used the lymph node metastases of interest which were seen in the radiotherapy planning CT scans. For treatment response the median follow-up period was 25 months (range 3-37). At the time of the last follow-up, 2 patients were dead for non-tumor related reasons. All patients showed initial response to radiotherapy. Two patients showed local progression at the primary tumor site after 15 and 19 months. There was one case of lymph nodal treatment failure after 12 months." [93]

8.1.3 Data acquisition and segmentation

"Each study participant underwent [MSOT] measurements of five target organs before, during (after 9 RT fractions), and approximately 90 days after RT treatment on both the right and left side of the neck. Target organs included the sternocleidomastoid muscle, the parotid, the submandibular and the thyroid glands, and, if possible suspect malignant or benign lymph nodes. A graphical overview of the measuring procedure is displayed in Figure 8.2. Each scan (static snapshot) of every target organ comprised three technical replicates. For the healthy cohort, the three temporal measurements of the five target organs were performed on the same day consecutively. For acquisition eleven wavelengths (680, 700, 730, 760, 800, 850, 920, 1000, 1030, 1064, 1100 nm) were used and the image processing was conducted using the Simulation and Image Processing for Photonics and Acoustics (SIMPA) toolkit, involving correction for wavelength-dependent laser power, [...] bandpass filtering, and four frames average per wavelength, followed by delay and sum beamforming, envelope detection (Hilbert transform), and linear spectral unmixing for Hb and HbO_2 using five wavelengths (700-850 nm). Clinical experts performed the semantic annotation of co-registered US images with the Medical Imaging Interaction Toolkit (MITK) [209] and the matching

of the lymph nodes over time using CT/MRI images acquired in the standard clinical workflow. " [93]

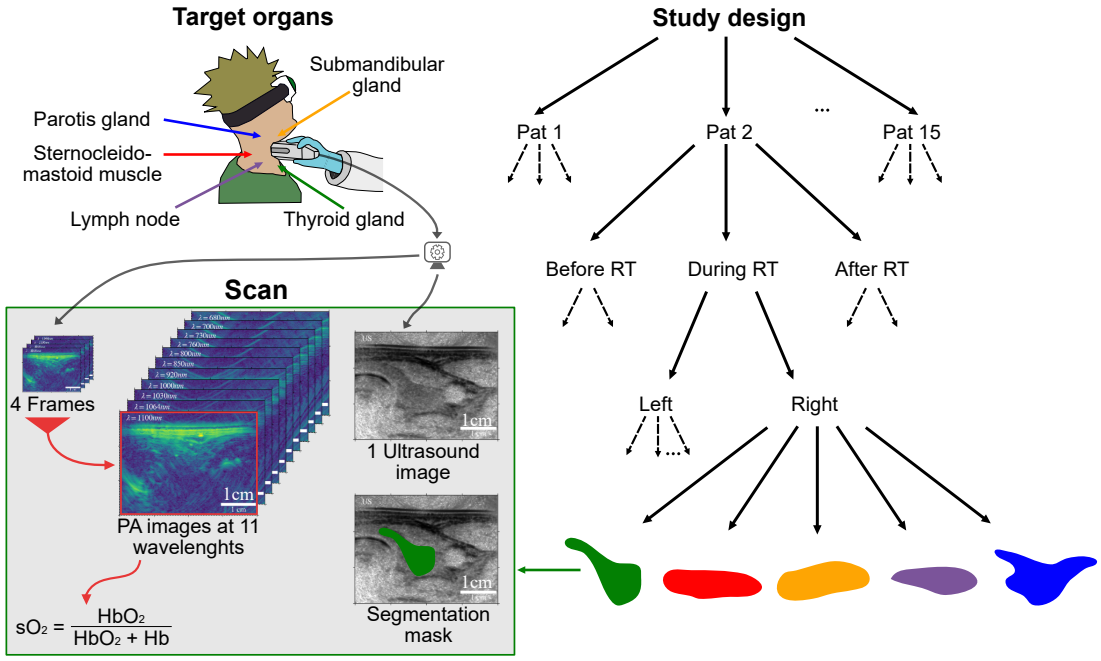


Figure 8.2: "The study involved photoacoustic (PA) measurements of five target structures (sternocleidomastoid muscle, parotid gland, submandibular gland, thyroid gland, and if detectable benign or suspect malignant lymph nodes) measured before, during (after the 9th fraction), and 90 days after radiotherapy (RT). For each patient (Pat) and measurement, multispectral photoacoustic images, co-registered with an ultrasound image were taken from the left and right side. Clinical experts labeled the ultrasound images to generate semantic segmentation masks. Functional parameters like oxygen saturation (sO_2) were derived from the multispectral information using linear spectral unmixed concentrations of deoxygenated (Hb) and oxygenated (HbO_2) hemoglobin." [93] Figure taken from [93] under CC BY 4.0 [120].

8.1.4 Linear mixed model analysis

To analyze the variance of the MSOT signal, a LMM analysis was performed. "The analysis of the explained variation was performed using the variance component form of the LMM [(Part I Sec.2.3.2)]. More specifically, we utilized separate linear mixed models for each wavelength to analyze explained variation and thus to evaluate the relevance (or importance) of factors contributing to changes in the observed absorbance

values (Fig. 8.6). More precisely, linear mixed models were fitted for each wavelength separately, with fixed effects for the time/treatment, location (side), depth (of the structure), number of pixels in the region of interest (counts), smoking history, body mass index (BMI), and season of measurement (WS). Random effects were employed for the factor subject as well as the unexplained (residual) variation (Eq. 8.1). The proportion of explained variance was derived through the empirical decomposition of explained variation based on the variance components version of the mixed model [215].

$$\begin{aligned}\text{absorbance}_{ij} = \alpha + \text{time}_{ij}^{\top} \cdot \beta_1 + \text{side}_{ij}^{\top} \cdot \beta_2 + \text{depth}_{ij}^{\top} \cdot \beta_3 + \text{counts}_{ij}^{\top} \cdot \beta_4 \\ + \text{smoking}_{ij}^{\top} \cdot \beta_5 + \text{BMI}_{ij}^{\top} \cdot \beta_6 + \text{WS}_{ij}^{\top} \cdot \beta_7 + \delta_i + \varepsilon_{ij},\end{aligned} \tag{8.1}$$

for repetition $j = 1, \ldots, n_i$ of patients $i = 1, \ldots, 12$. The number of repetitions n_i varied per patient. Here, α denotes a fixed intercept, and the β_k, $k = 1, \ldots, 6$ denote the effect sizes of the corresponding fixed effect. The random intercept $\delta_i \sim \mathcal{N}(0, \sigma_\delta^2)$ describes subject specific variation. The residuals $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ capture the residual or unexplained variation. Within the model, we assume that the random effects and the residuals are stochastically independent. Similarly, we used a single LMM to investigate the relevance of variables on the oxygenation of patients. "[93]

8.2 Results

The following section presents the results of PAI measurements structured around three primary clinical hypotheses: (1) an early increase in tumor sO_2 may indicate a positive treatment response, (2) a late increase in sO_2 in surrounding tissue may suggest inflammation, and (3) a late decrease in water content may reflect reduced salivary gland function. These analyses are followed by a brief qualitative outlook on organ-specific spectral fingerprints.

8.2.1 Measured oxygenation does not reflect prior knowledge

"According to the literature (e.g. Steel et al. 1989 [216]), we anticipated reoxygenation of malignant lymph nodes resulting from RT treatment (H1). In contrast to this expectation, our PAI measurements indicated a discernible decline in the estimation of the biomarker sO_2 within most of the malignant target lymph nodes as a result of RT (Fig. 8.3). In particular, we observed a strong increase in deoxyhemoglobin

levels during and after RT, without a pronounced trend for oxygenated hemoglobin ([Appendix] Figs. B1 and B2). For visual reference, images depicting a malignant lymph node throughout RT are presented in the [Appendix] Figure B3." [93] Additionally, two lymph node PAI measurements series, one exhibiting tumor progression and one remaining progression-free, are presented as exemplary case studies in Appendix Figures B4 and B5, respectively.

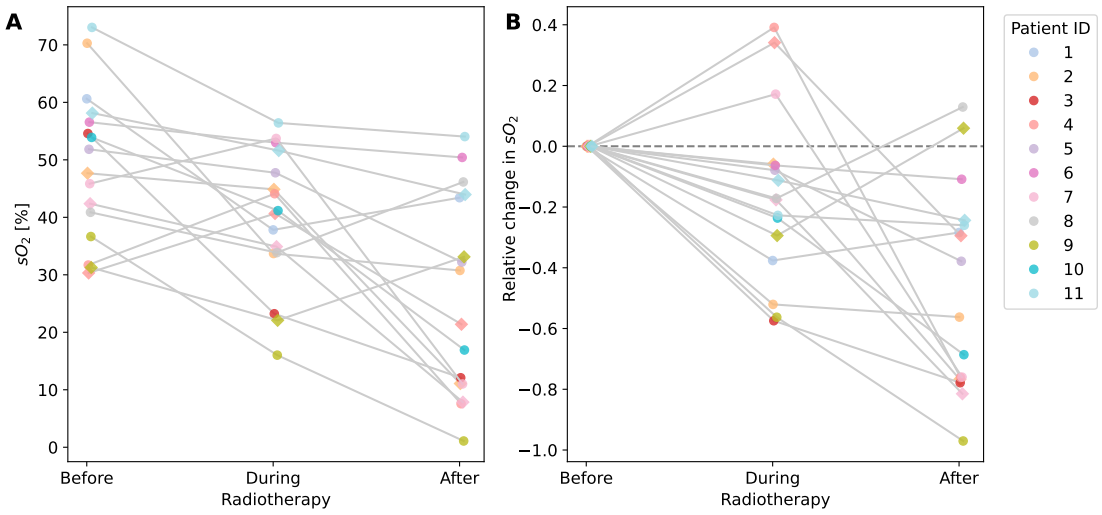


Figure 8.3: "Oxygen saturation (sO_2) before, during (after the 9th fraction), and 90 days after radiotherapy (RT) treatment in suspect lymph nodes, as measured by photoacoustic imaging (PAI), decreases over time. This is shown for the overall sO_2 values (A) and relative to the pre-RT sO_2 (B), with each node color-coded by patient. For some patients, multiple nodes were measured, as indicated by their shape (1st node: circle; 2nd node: diamond shape)." [93] Figure taken from [93] under CC BY 4.0 [120].

8.2.2 Linear mixed model analysis reveals largely unexplained signal variability

"For better interpretability of our results, we performed an analysis of the explained variation using the variance component form of the linear mixed model (LMM) [215]. In an ideal imaging setting, most variability in measurement would be explained by therapy-induced differences in tissue structure and function such as reoxygenation, revascularization, or necrosis. However, real-life imaging is influenced by numerous additional factors contributing to the signal generation for example subject, depth of the organ of interest, and seasonal variations, potentially shadowing the actual target property. Our results indicate that the variance of sO_2 in our measurements [of

malignant lymph nodes] is largely unexplained (40%). Only 32% is treatment-induced (temporal variation) while 32% can be attributed to intra-subject variability ([Appendix] Tab. B6)." [93]

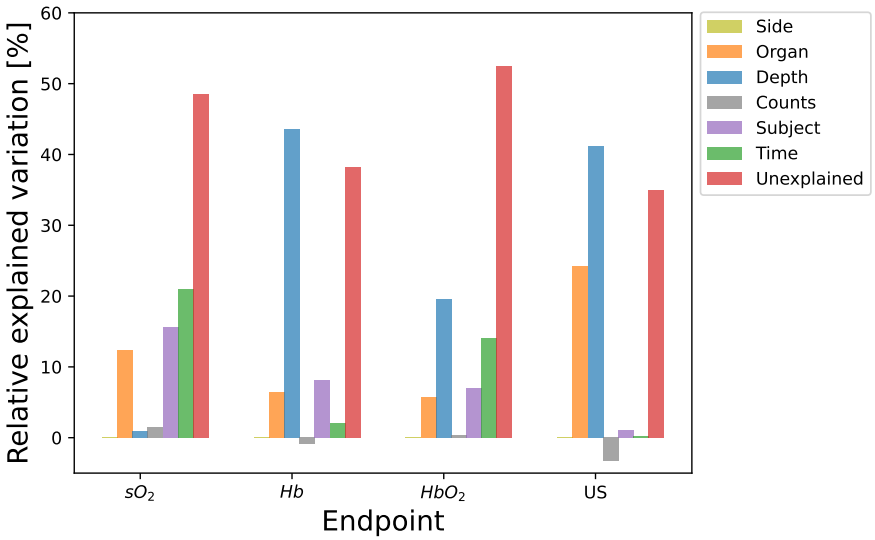


Figure 8.4: Variance decomposition for the endpoints oxygen saturation (sO_2), deoxygenated hemoglobin (Hb), oxygenated hemoglobin (HbO_2), and ultrasound (US), expressed as percentages, derived from the linear mixed model (LMM) analysis of all five measured organs. The variability in photoacoustic patient measurements is predominantly explained by unexplained residuals and imaging depth particularly for Hb and HbO_2. In contrast, depth has minimal influence on sO_2. Treatment time point (temporal variation), organ class, and intra-subject variation each contribute modestly to the overall variance across photoacoustic-derived endpoints. For the US signal, variance is particularly more organ-specific, with additional main contributions from depth-dependent and unexplained effects, while temporal and intra-subject variation are negligible.

In line with the previously presented LMM analysis focused on sO_2 within the suspicious lymph node class, an extended LMM analysis was performed on all organ data of the patient cohort, evaluating four endpoints: sO_2, Hb, HbO_2, and US (Fig. 8.4). The results indicate that imaging depth and residual (unexplained) variation account for the majority of variance in Hb and HbO_2. In contrast, depth contributes only minimally to variation in sO_2, reflecting its nature as a relative measure. Furthermore, treatment time point (temporal variation), intra-subject variability, and organ-specific effects each account for approximately 2%-14% of the variance in Hb and HbO_2, while

these same factors explain a more substantial proportion of the variance in sO_2.

In addition to the large proportion of unexplained and depth-related variance, the US endpoint also exhibits a strong organ-specific component, as expected. This highlights the potential of US for anatomical classification based on structural features. Notably, and in contrast to PAI-derived parameters, time (treatment effect) has no measurable effect on US signal variance, underscoring the fact that US does not capture functional or physiological information relevant to therapy response.

Residual variance attributable to hemoglobin level fluctuations Numerous factors can influence medium-term changes (over days to weeks) in hemoglobin concentration, including time of day [217], dietary intake, hydration status, and medication use. Given that hemoglobin is a major absorber and thus a primary contrast source in PAI, such physiological variations are likely to influence the stability and interpretability of the photoacoustic signal.

In routine clinical practice, blood sampling including tHb measurements are standard components of RT monitoring. However, these blood samples were not consistently time-matched with PAI acquisitions, limiting their immediate interpretive value. To assess the relationship retrospectively, a post hoc analysis was performed comparing blood sample-derived tHb levels with the corresponding MSOT-derived tHb estimates, computed as the mean tHb value across all imaged organs per patient (Appendix Fig. B6). The analysis revealed a moderate correlation (Spearman coefficient) of 0.48 (95% CI: [–0.07, 0.81]; p = 0.08), under the assumption that each data pair (PAI tHb – blood tHb) was independent.

However, this independence assumption is clearly an oversimplification, as supported by the substantial intra-patient explained variance observed in the LMM analysis (Fig. 8.4). To account for this, a more conservative analysis was performed using only one pair per patient (limited to the ten patients with matched data prior to the start of RT). In this subset, the correlation remained at 0.48, with a wider confidence interval (95% CI: [–0.22, 0.85]) and reduced significance (p = 0.16), reflecting the small sample size (Appendix Fig. B7). CIs were computed using Fisher's z' transformation, assuming a bivariate normal distribution (Part I Sec. 2.3.1) [77].

8.2.3 Photoacoustic imaging suggests signs of inflammation in the sternocleidomastoid muscle after radiotherapy

"Current RT treatments are becoming progressively more and more targeted. Nonetheless, a considerable portion of the radiation is still delivered in organs surrounding the primary target (here: lymph nodes) resulting in local inflammation [218]. In support of H2, we observed an increase in tHb concentration in the sternocleidomastoid muscle, especially at the last (third) measurement (Fig. 8.5). The tHb estimates before and during RT, on the other hand, are in a similar signal range to the measurements of the healthy cohort (Fig. 8.5 B). The observed increase in tHb is mainly due to an increase in deoxyhemoglobin, which approximately doubled between the second and third measurements indicating tissue activity ([Appendix] Fig. B8). This, coupled with a drop in oxygenated hemoglobin, results in a decline in derived sO_2.

It is a well-known fact that the functional tissue parameters provided by the PAI system can only serve as rough approximations of the ground truth, primarily due to the spectral coloring effect in tissue [40]. To overcome this issue, we investigated the tumor-to-background ratio, which is a common measure to derive radiopharmaceutical uptake using PET [(positron emission tomography)] [219, 220]. To differentiate between surrounding and tumor tissue response to radiation we calculated the tumor-to-muscle ratio (TMR) for sO_2. This ratio exhibits an increasing trend ([Appendix] Fig. B9), primarily driven by a strong increase in deoxyhemoglobin under RT of the sternocleidomastoid muscle. In contrast to the raw sO_2 values, which are heavily affected by spectral coloring, the normalized TMR values are indicating reoxygenation of the lymph nodes in line with H1." [93]

8.2.4 Photoacoustic imaging reveals changes in water content indicating xerostomia

"In addition to providing estimates of hemoglobin concentrations and hence sO_2, multispectral PAI promises to retrieve a diverse array of absorbers present in the near-infrared, enabling the investigation of a wide range of biological effects. Specifically, water emerges as a relevant chromophore for wavelengths around and above 1000 nm (absorption spectra Fig. 8.6 B - background). We suggest linking water concentration to a prevalent RT side effect, xerostomia, characterized by patients suffering from dry mouth.

The water content before and during the RT is approximately constant and comparable to the healthy volunteer group ([Appendix] Fig. B10). Our measurements after RT indicate a decrease in water content visible in the PAI signal of the primary

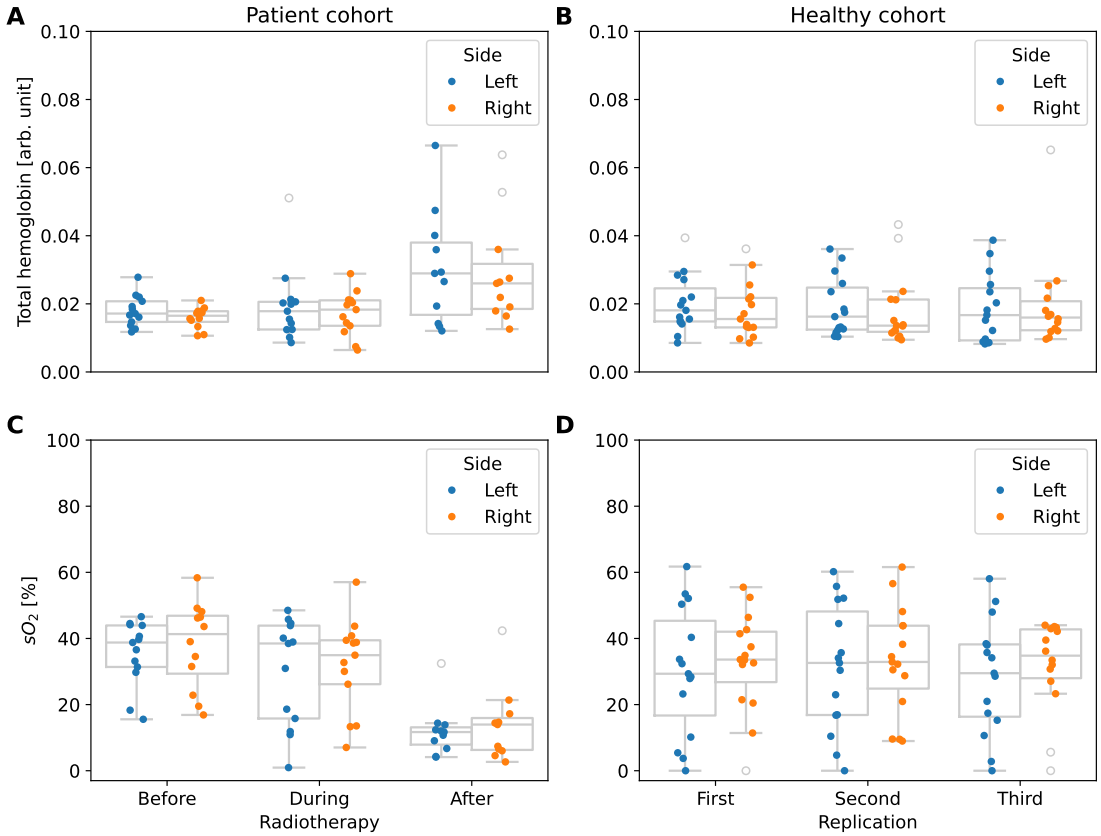


Figure 8.5: "Photoacoustic imaging indicates evidence of inflammation in the sternocleidomastoid muscle following radiotherapy (RT) treatment, characterized by an increase in total hemoglobin (A). The increase is mainly caused by deoxyhemoglobin ([Appendix] Fig. B8) leading to a drop in oxygen saturation (sO_2) (C). Healthy volunteer data (no RT - three temporal replications at one session) are displayed for comparison (B/D). The box shows the interquartile range (IQR) with the median as center line. The whiskers extend to points that lie within 1.5 * IQR of the 1st or 3rd quantile. All data points are plotted on top, and color-coded by measurement side (left or right neck). Outliers are shown without color." [93] Figure taken from [93] under CC BY 4.0 [120].

salivary gland (Fig. 8.6 A), the submandibular gland, which is responsible for about 60% of saliva production in the unstimulated state [221]. This observation is supported by an LMM analysis for the PAI signal of the 11 measured wavelengths within the submandibular gland. The LMM analysis (Fig. 8.6 B) shows an increase in explained variation of the variable treatment with longer wavelengths. This goes alongside an

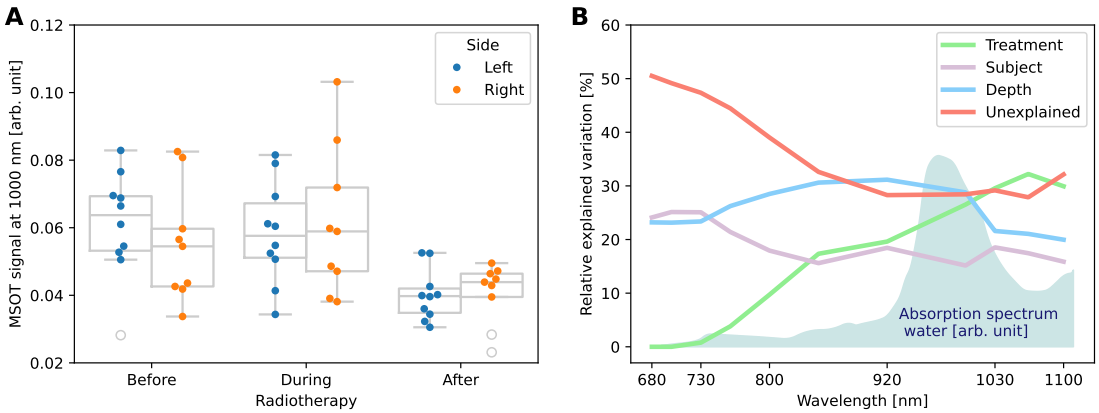


Figure 8.6: "The photoacoustic signal of the submandibular gland at 1000 nm shows a strong correlation to the treatment phase (A). In line with this finding, the linear mixed model (LMM) analysis reveals an increase in explained photoacoustic signal variation for the treatment parameter (green) with longer wavelengths, where water becomes a primary absorber (B). Other LMM parameters like the individual (subject) and the depth of the structure stay constant or decrease with longer wavelengths like the unexplained (residual) signal variation. The box shows the interquartile range (IQR) with the median as center line. The whiskers extend to points that lie within 1.5 * IQR of the 1st or 3rd quantile. All data points are plotted on top, and color-coded by measurement side (left or right neck)." [93] Figure taken from [93] under CC BY 4.0 [120].

increase of water absorption towards longer wavelengths and by declining inter-subject, depth, and particularly unexplained variability (residual variance)." [93]

8.2.5 Organ fingerprints

As different organs exhibit distinct molecular compositions and biological functions, PA measurements should, in principle, encode organ-specific spectral information. In other words, each organ is expected to exhibit a unique spectral fingerprint. Such fingerprints could be valuable for applications such as semantic segmentation, aiding the automated and standardized analysis of study data, or for analyzing treatment-induced spectral changes.

Supporting this hypothesis, the LMM analysis revealed that organ identity contributes to the explained variance in the PA signal, reinforcing the idea that measurable organ-specific spectral characteristics exist. A direct analysis of organ-wise signals showed that the mean spectra across organs exhibit visually distinct features (Fig. 8.7).

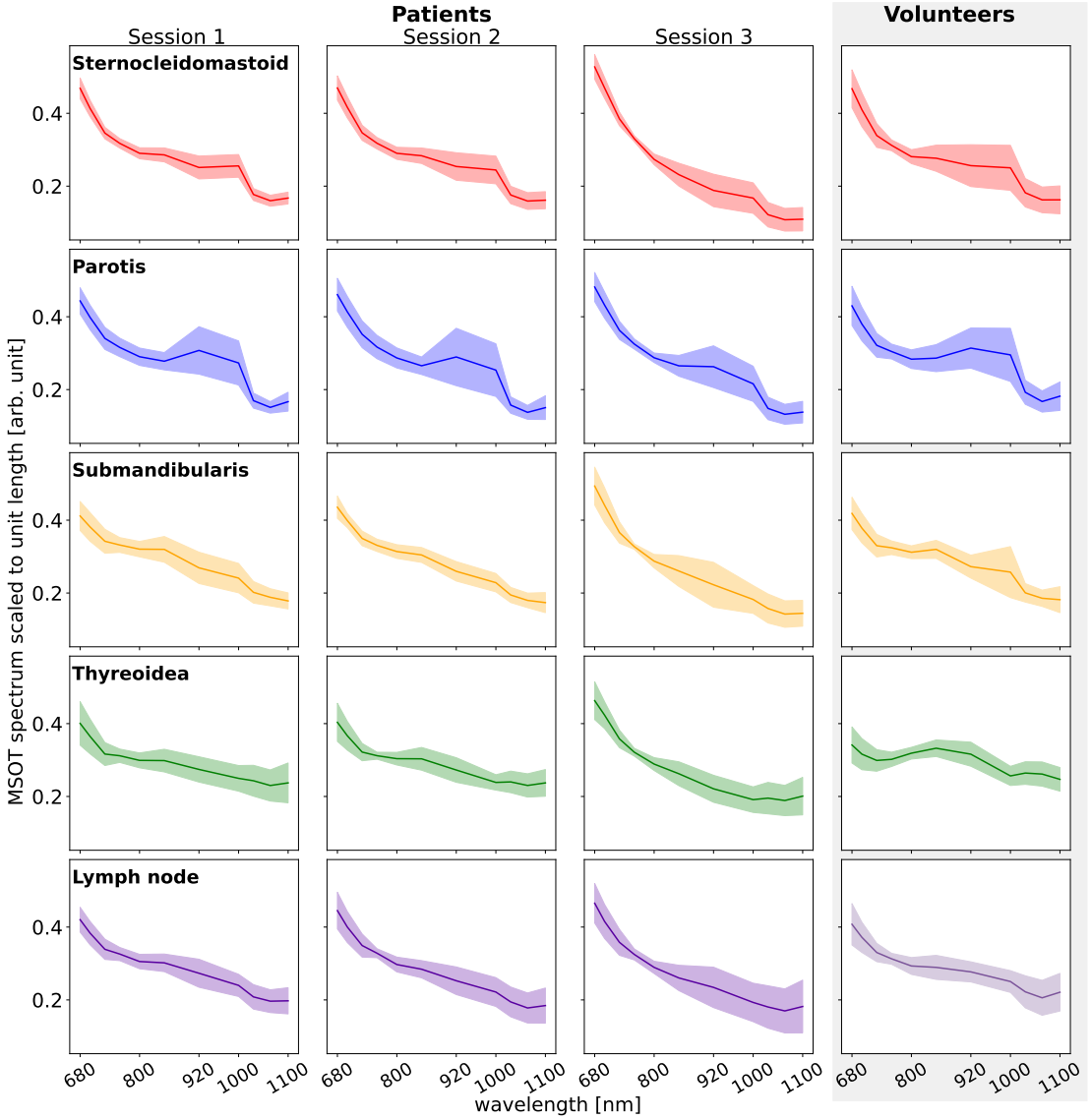


Figure 8.7: Mean spectra with standard deviation normalized to unit vector length of photoacoustic (MSOT) signal for five target anatomical structures, plotted from top to bottom: sternocleidomastoid muscle, parotid gland, submandibular gland, thyroid gland, and lymph nodes with light purple representing benign and dark purple representing suspect malignant nodes. From left to right, spectra are shown for three imaging sessions (before, after the 9th fraction, and 90 days after radiotherapy) followed by spectra from healthy volunteers.

These differences become even more prominent when examining the spectral gradient (Appendix Fig. B11), defined as:

$$\text{Spectral Gradient} = \frac{\Delta S}{\Delta \lambda} = \frac{S_{i+1} - S_i}{\lambda_{i+1} - \lambda_i}, \tag{8.2}$$

where S_i and S_{i+1} are the PA signal values at wavelengths λ_i and λ_{i+1}, respectively. Further, a supervised dimensionality reduction approach using UMAP [222] (Appendix Fig. B12) demonstrated that it is principally possible to separate organs based on their spectral characteristics, yielding a low-dimensional embedding of high-dimensional spectral data.

However, attempts to apply this approach to unseen data failed to generalize, indicating that robust fingerprinting across individuals or imaging conditions remains challenging. These findings are consistent with results from a master's thesis [223] by Patricia Vieten, which I co-supervised and presented as poster at the *SPIE BiOS* 2022 [224]. In her work she investigated whether deep learning-based multilabel semantic segmentation of HN ultrasound images could benefit from the addition of PA information. Using the nnU-Net ('no-new-Net') segmentation framework [225] and data from the 15 healthy volunteers, the study found that multispectral PA data did not improve segmentation performance beyond that achieved with ultrasound alone. Furthermore, it was shown that single-pixel PA spectra of deeper anatomical structures lacked characteristic features suitable for classification or clustering. Supporting simulations using in silico digital twin models within the SIMPA framework further revealed that optical fluence attenuation substantially alters the spectral profiles of deeper tissues, likely limiting the effectiveness of fingerprinting approaches without compensation strategies.

8.3 Discussion

"This work presents the first in-human PAI study for monitoring the response of malignant lymph nodes in head and neck cancer patients to RT over time. Our study revealed radiation-induced changes in the surrounding tissue, including an increase in derived total hemoglobin in the sternocleidomastoid muscle, indicative of inflammation and revascularization, and a decrease in water content in salivary glands, suggesting xerostomia, confirming our hypotheses H2 and H3. Strikingly, PAI is capable of measuring molecular changes induced by RT in human tissue non-invasively. We also observed an increase in total hemoglobin within malignant lymph nodes, primarily comprising deoxygenated hemoglobin, resulting in a decline in sO_2. This

decrease in sO_2 contradicts our initial hypothesis (H1) of observing reoxygenation within malignant lymph nodes resulting from RT treatment, as summarized in [216]. A digital twin model shed light on PAI-specific challenges in signal generation from malignant lymph nodes, particularly revealing low BVF as a factor associated with high prediction errors in sO_2 [(Part II Sec. 7)]. Our initial findings highlight the potential of PAI in directly monitoring tissue responses under RT at a biological scale, particularly inflammation, revascularization, and tissue malfunction, such as xerostomia, which has the potential to expand current clinical capabilities. To achieve these benefits PAI must overcome several barriers [6]. Notably, the limitations of photoacoustics, as described by Noltes et al. (2023) [10], including acoustic reflections, penetration depth (fluence and spectral coloring), spectral cross-talk, and strong absorber coverage, are prominent in the data presented in this study. Additionally, reoxygenation effects predominantly occur in the hypoxic regions deep inside the lymph nodes, where fluence and coloring effects are most pronounced. As a treatment outcome, the nodes undergo shrinkage, posing challenges in matching individual nodes over time in particular for the last measurement. While a recent study shows the stability of PAI signal levels in healthy volunteers over 14 days [226], anatomical changes due to the shrinkage of the lymph nodes, amongst others, limit the direct comparisons of PAI signals throughout the present study. Therefore, we also investigated relative measures such as the TMR and the explained variation with LMMs, which indicates high residual variance. Moreover, spectral coloring effects are particularly influential in shrinking nodes, as they often settle deeper in tissue post-shrinkage. The increase in tHb in surrounding tissue enhances the coloring effect. While the derived tumor-to-muscle ratio could potentially aid in identifying this phenomenon, this normalization should be interpreted cautiously to avoid confirmation bias and should be verified in future studies. A post hoc analysis of independently taken blood samples from the patients revealed a moderate correlation between tHb levels derived from the blood samples and PAI-derived tHb concentration within the target organs (Spearman correlation coefficient of 0.48 with p=0.08). Therefore, blood tHb levels may pose as a potential confounder, given the moderate correlation to the PAI signal, necessitating further investigation." [93] However, these results must be interpreted with caution. The observed correlation is based on a composite signal integrating data from multiple organs, each undergoing different biological responses to radiotherapy, which may obscure organ-specific relationships. A more targeted follow-up study in healthy volunteers, controlling for organ-specific and without treatment effects, is therefore warranted to validate and isolate the influence of blood tHb levels on PAI signal fidelity. "Furthermore, malignant lymph nodes pose

an aggravating factor due to their low BVF and high acoustic attenuation, resulting in particularly weak measurable photoacoustic signals." [93] Additionally, although spectral fingerprint characteristics appear to be present, the inclusion of functional photoacoustic information in ultrasound-based semantic segmentation using the nnU-Net architecture did not lead to performance improvements. This outcome is likely attributable, at least in part, to fluence-related distortions, as suggested by supporting evidence from digital twin simulations. However, "besides endeavors within the PAI community to tackle the quantification problem, identifying the most relevant feature(s) to address clinical problems is crucial. Therefore, we propose to further investigate tHb and relative measures as the TMR for monitoring RT response, since the interplay between oxy- and deoxyhemoglobin and current unmixing techniques might add more uncertainty in the low BVF regime of malignant lymph nodes (see digital twin analysis). Although oxygenation is highly relevant for assessing RT treatment response and tumor status, ambiguous measures for differences in sO_2 of malignant and benign thyroid nodules have been reported in recent PAI literature. Roll et al. [227] observed significantly lower sO_2 in malignant than benign nodes and Noltes et al. [10] "could not determine statistically relevant differences between benign and malignant thyroid nodules based on mean oxygen saturation in thyroid nodules". Additionally, water content could be explored for assessing xerostomia. On the other hand, we see the importance of disentangling therapeutic effects from individual and PAI-specific uncertainties. This could be achieved by:

- Collecting larger datasets to understand heterogeneities in treatment response better.
- Implementing a meaningful normalization, such as the tumor-to-muscle ratio, to separate the therapeutic effect from individual variations.
- Identifying and addressing confounding factors that may influence PAI measurements.

Addressing these aspects may enhance the clinical utility of PAI for monitoring RT response and improving patient outcomes." [93] Additionally, analyzing subvolumes as demonstrated by Zips et al. [228] using hypoxia specific PET imaging, may help to mitigate unwanted variability, particularly if focused on the initial treatment time points, where anatomical changes are minimal.

"In conclusion, the heterogeneity within the disease and the substantial signal variability arising from the current limitations of PAI highlight the challenges in monitoring

RT treatment of malignant lymph nodes in head and neck cancer. Despite these obstacles, PAI shows promise as an early non-invasive assessment method for radiation response and could become a valuable tool in clinical practice, enhancing patient care and outcomes in head and neck cancer treatment." [93]

9 Diagnosis of peripheral artery disease

Despite the availability of various non-invasive diagnostic options for peripheral artery disease (PAD), such as ABI, patients presenting with unclear leg pain during walking remain challenging to diagnose [114], often necessitating invasive and costly angiography. In these cases, a diagnostic tool that allows for the direct assessment of the affected organ, namely the lower leg musculature, would be highly valuable.

However, as shown in the previous section, successfully resolving tissue oxygenation and detecting changes in tissue with 2D PAI remains a difficult task. Therefore, this study aims at the first in vivo application of the optical pattern approach for 3D PAI to diagnose PAD. The analysis is based on longitudinal pattern measurements taken before (pre) and after (post) athletic exercise, followed by subsequent temporal volume registration (Part II Sec. 5.1.2.4) for both a PAD and a healthy volunteer (HV) group.

This project was conducted in collaboration with the department of vascular surgery, University Hospital Erlangen, Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg, Erlangen, mainly involving Ulrich Rother, Josephine Günther, Marcella Staus, and Yi Li.

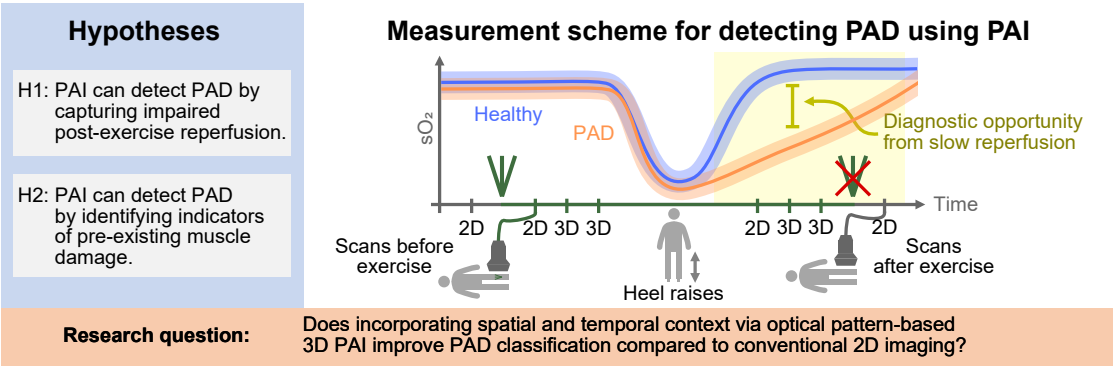


Figure 9.1: Mission of this work: to investigate whether photoacoustic imaging (PAI) can detect pathophysiological changes in muscle tissue caused by peripheral artery disease (PAD). The specific focus was to evaluate whether impaired post-exercise reperfusion (H1) and pre-existing muscle damage (H2) manifest in PA measurements, and whether incorporating spatial and temporal context via optical pattern-based 3D imaging improves classification performance compared to conventional 2D approaches.

PAD classification The fundamental assumption in classifying PAD using PA images is based on the hypothesis that individuals with PAD exhibit impaired perfusion

or delayed reperfusion of the muscle tissue when compared to HVs, particularly following athletic exercise (Fig. 9.1). Additionally, continuous malperfusion in PAD is expected to result in structurally and functionally heterogeneous, and potentially damaged, muscle tissue [112]. Such alterations should be detectable in both pre- and post-exercise states. This motivates the following working hypotheses (Fig. 9.1):

H1: PAI can detect PAD by capturing impaired post-exercise reperfusion.

H2: PAI can detect PAD by identifying indicators of pre-exercise muscle damage.

These two hypotheses are examined as part of the overarching research question: Does incorporating spatial and temporal *context* via optical pattern-based 3D PAI improve PAD classification compared to conventional 2D imaging?

9.1 Materials and methods

This section provides a detailed description of the study design and the key evaluation methods applied. The study was conducted in two parts: an initial pilot study [229] and a subsequent follow-up study. The primary objectives of the pilot study were to establish a standardized measurement protocol and to perform case number calculations to determine the required sample size for the follow-up study. The follow-up study included a larger cohort. This part of the study was designed to systematically investigate the application of the developed methodology and to enable robust statistical analysis.

9.1.1 Ethics

These studies were conducted in accordance with the guidelines of the Helsinki Declaration, were approved by the ethics committee of the medical faculty of the FAU, and were registered at clinicaltrials.gov (ID pilot: NCT05110677; ID follow-up: NCT05773534). All study procedures were carried out with the full informed consent of all participants.

9.1.2 Study design, data acquisition and processing

Pilot study The pilot study included N=9 individuals, consisting of 5 PAD patients and 4 HVs. The PAD group covered Fontaine classification stages (Part I Sec. 2.4.2) from stage II (claudicatio) to stage IV. PAD classification was based on clinical reference standards, including parameters such as angiography and ABI measurements, where an ABI value of less than 1 was indicative of PAD.

Each participant underwent four MSOT measurements of the calf before and after an athletic exercise (walking 150 m). The measurement sequence was structured as follows: [2D, pattern attachment, 2D, 3D, 3D, exercise, 2D, 3D, 3D, 2D] (Fig. 9.1). For the 3D measurements, the preset included the wavelengths 665, 730, 760, 800, 665, 850, 930, and 1030 nm. For the 2D measurements, a preset covering 700 – 950 nm in 10 nm steps was used.

The optical pattern measurement and data processing were performed as described in Part II Section 5.1.3.2, using the first inkjet-printed iteration pattern with a trident width of 30 mm and a length of 50 mm.

Based on the sO_2 measurements obtained from the pilot study, case number calculations were performed by Dr. Silvia Calderazzo from the division of biostatistics at DKFZ. The analysis recommended acquiring at least N=15 HVs and N=15 PAD patients to achieve a statistical power of 0.8 (probability of detecting a true effect, indicating that either 2D or 3D imaging improves PAD classification) with an α value of 0.05 (probability of incorrectly rejecting the null hypothesis).

Follow-up study Based on novel insights from the pilot study, specifically that preset changes between the 2D and 3D measurements frequently led to system breakdowns, along with a clinical focus shift toward claudicatio and preliminary results from a parallel clinical study by our collaborators [177], which showed that heel raises induced a more intense workout, the protocol for the follow-up study was accordingly adapted, rendering the pilot data unsuitable for inclusion in the joint analysis.

In the follow-up study, an initial cohort of N = 33 individuals was measured in early 2023. To further improve the sex-related balance, an additional N = 13 individuals were recruited in early 2024. The resulting dataset of N = 46 individuals comprises 22 PAD patients (exclusively Fontaine stage II) and 24 HVs. The overall measurement protocol remained consistent, with the exception that only the 3D preset was used, and heel raises replaced the initial walking exercise as the standardized athletic intervention. Post-measurement analyses revealed that between the first cohort in 2023 and the second cohort in 2024, maintenance was performed, including replacing the probe membrane, which noticeably influenced the PAI signal and may represent a confounding factor. Additionally, and without our knowledge, the order of the 3D preset was changed to 665, 665, 730, 760, 800, 850, 930, and 1030 nm.

For this follow-up, the screen printing iteration pattern (trident width = 30 mm, length = 50 mm) was used for all participants except the first 6 patients, who were measured using a screen printed N-shaped pattern (length = 55 mm, width = 25 mm).

As described in Part II Section 5.3, and shown in Appendix Fig. A8, the rotational angle α cannot be recovered from the N-shaped geometry. Due to the stable acquisition (Part II 5.2.1.1), it was decided to include these data in the evaluation by assuming α=0.

Datasets The data were processed according to the protocol described in Part II Section 5.1.3. This processing resulted in three distinct datasets: the static 2D dataset, the optical pattern-based compounded 3D dataset, and an intermediate naïve 2D image stack referred to as the sweep dataset. The sweep dataset consists of all measured 2D frames from the 3D acquisition but without compounding-based spatial alignment.

To assess the classification performance, the dataset was divided into training and test sets using a 70/30 split (31/15 individuals). The allocation was carefully balanced with respect to health status, sex, and measurement time point (before/after maintenance) to ensure an unbiased and representative distribution (Fig. 9.2). The test set remains untouched throughout this thesis due to ongoing analyses.

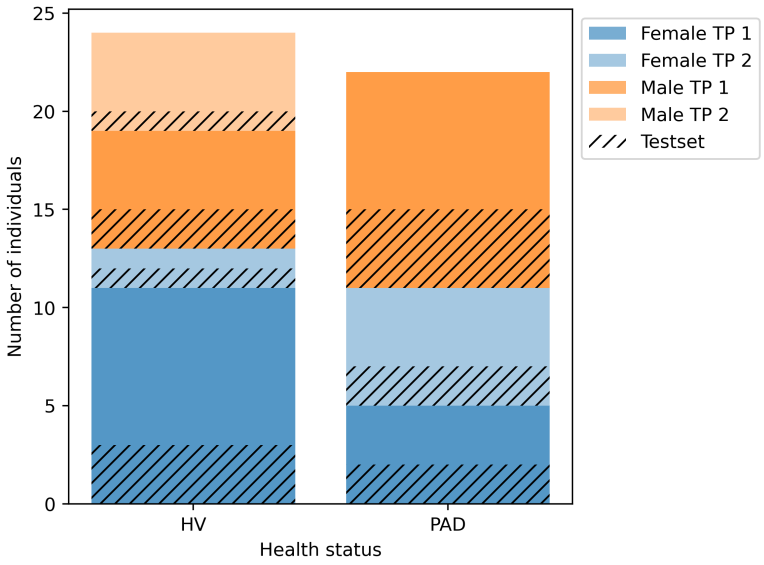


Figure 9.2: Data split of the peripheral artery disease (PAD) follow-up study. Sex is indicated by color (blue: male, orange: female), while the color intensity represents the measurement time point (TP), with intense colors indicating pre-maintenance and pale colors indicating post-maintenance measurements. Hatched bars indicate data reserved for the test set. HV: healthy volunteer.

Features For each frame, an ROI measuring 20 mm $\times$ 5 mm and located 3 mm below the skin surface was selected. Within this ROI, the mean signal intensity of the

PA images at each wavelength, as well as that of the spectrally unmixed Hb, HbO_2, and sO_2 images (obtained via linear spectral unmixing using 730-850 nm), was extracted per frame and used as a feature for further analysis. The PA spectrum within this ROI was normalized to unit length after unmixing. To evaluate the classification performance, the metrics recommended by the metrics reloaded toolkit [68] were used, specifically accuracy, LR^+, and AUROC, as detailed in Part I Section 2.3. A summary of the metric selection process is provided in Appendix Table C1.

Classification For classification, a random forest classifier from `sklearn.ensemble` was employed as described in Part I Section 2.3.2. The classifier parameters are provided in Appendix Table C2. To maximize the data available for validation, a leave-one-out (LOO) cross-validation scheme was applied, training a total of 31 models. Each model was trained on data from 30 individuals and validated on the remaining individual to enable framewise classification of PAD.

The framewise predictions were aggregated by calculating the mean of the prediction labels to obtain a final classification result for each individual. This approach yielded 31 individual predictions used for evaluation. Unless otherwise stated, all features were used in the classification: [730, 760, 800, 850, 930, 1030, Hb, sO_2, HbO_2].

Feature combination To assess the functional response to exercise, both pre- and post-exercise data were used as input features. The framewise matching of pre- and post-exercise data was performed based on frame indices. For the 2D dataset, this matching lacks any spatial correspondence or meaningful spatial alignment. For the sweep dataset, the matching was semi-paired since low frame indices correspond to measurements near the pattern origin, whereas higher indices represent positions farther along the pattern. For the optical pattern-based 3D reconstructed dataset, frame indices directly correspond to defined distances from the pattern origin, yielding a fully paired dataset. Additionally, a variant of the 3D dataset was created using the temporal volume registration method described in Part II Section 5.1.2.4, in which alignment between pre- and post-measurements was refined using rigid registration.

9.2 Results

In this section, the experiments and results of the first application of the optical pattern-based approach for 3D PAI in humans are presented. The primary aim is to demonstrate the diagnostic potential of PAI, highlighting the advantages of 3D PAI

over conventional 2D imaging for classifying PAD by leveraging perfusion dynamics and muscle heterogeneity.

9.2.1 Data exploration reveals strong sex separability

A PCA performed on the mean spectrum from a 20 mm x 5 mm cropped ROI 3 mm below the skin across all functional wavelengths (730, 760, 800, 850, 930, 1030 nm) using all measured frames of the training set reveals a strong intra-subject correlation, and additionally an inter-subject separability (Fig. 9.3 **A**). While neither health status (Fig. 9.3 **B**) nor the measurement frames taken before and after exercise (Appendix Fig. C1) exhibit distinct clustering patterns, sex appears to form two clusters (Fig. 9.3 **C**). These clusters are still close and partially overlapping but are distinguishable in the PCA space.

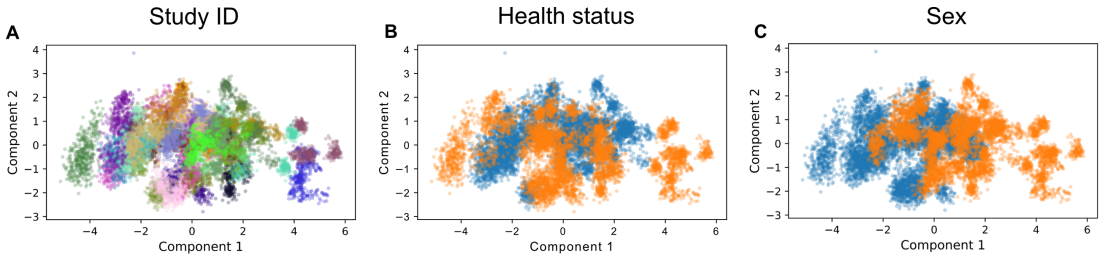
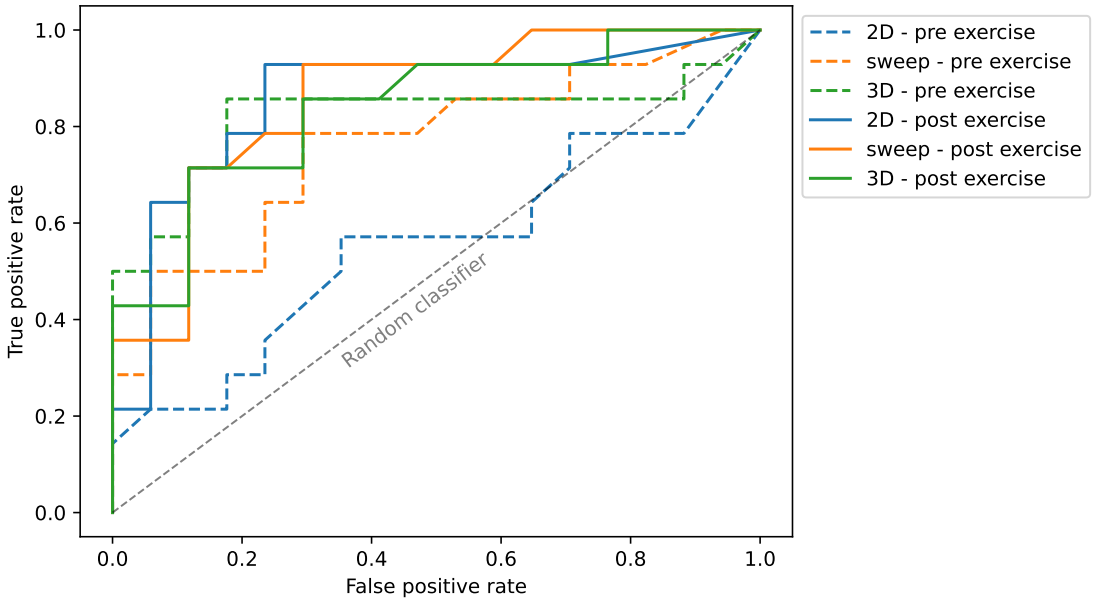


Figure 9.3: Principal component analysis (PCA) (`sklearn.decomposition`) of the training set from the peripheral artery disease (PAD) follow-up study. Each dot represents the derived principal components, calculated from the mean value within a 20 mm x 5 mm cropped region of interest of one frame of one scan across all functional wavelengths (730, 760, 800, 850, 930, 1030 nm). **A** Shows color coding by study ID, **B** shows color coding by health status (blue: healthy, orange: PAD), and **C** shows color coding by sex (blue: female, orange: male). The explained variance ratio for the PCA components can be found in Part I Figure 2.8.

9.2.2 Optical pattern-based 3D imaging is beneficial for pre-exercise PAD detection

The most predictive single parameter for classifying PAD was the mean HbO_2 measured post-exercise, analyzed using multilevel thresholding. This parameter achieved an AUROC of approximately 0.76 [95% CI 0.6–0.9] across all three datasets when aggregating all post-exercise frames per study. Particularly sO_2 and 850nm absorption, yielded higher classification performance than other parameters (Appendix Tab. C3).



dataset	AUROC	95% CI	accuracy	95% CI	LR+	95% CI
2D-pre	0.55	[0.34, 0.77]	0.58	[0.39, 0.77]	1.42	[0.59, 4.7]
sweep-pre	0.76	[0.59, 0.94]	0.68	[0.52, 0.84]	2.43	[0.94, 10.91]
3D-pre	0.82	[0.64, 1.0]	0.81	[0.68, 0.94]	6.07	[2.31, 'n/a']
2D-post	0.86	[0.71, 1.0]	0.77	[0.61, 0.9]	5.46	[1.87, 'n/a']
sweep-post	0.86	[0.72, 0.99]	0.81	[0.68, 0.94]	6.07	[2.25, 'n/a']
3D-post	0.84	[0.69, 0.98]	0.77	[0.61, 0.9]	5.46	[1.9, 'n/a']

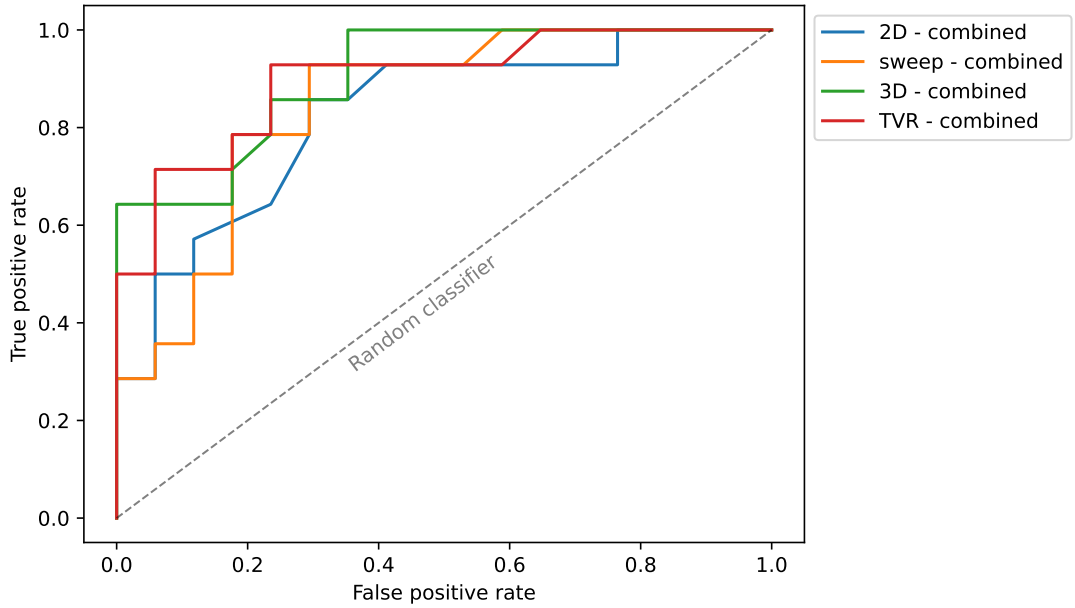
Figure 9.4: Peripheral artery disease (PAD) classification using a random forest on the static 2D, sweep, and 3D datasets pre- and post-athletic exercise shows little difference in classification results post-exercise. Only the 2D pre-exercise classification performs approximately at the level of random guessing. LR^+ 'n/a' corresponds to 100% specificity (Eq. 2.33).

Improved classification results were obtained using a random forest classifier and multiple features (Fig. 9.4). Post-exercise, the classification performance across all three datasets was fairly robust, exemplified by an AUROC of approximately 0.86 [0.7–1]. In contrast, the pre-exercise results painted a different picture. The 2D dataset performed essentially at the level of random guessing. The sweep dataset, however, showed a noticeable improvement with an AUROC of 0.76 [0.59–0.94], and the 3D dataset achieved even better classification results, comparable to those obtained post-exercise.

A two-sided paired t-test (`scipy.stats.ttest_rel`) on the prediction accuracy of

the pre-exercise datasets indicates that the 3D dataset significantly outperforms the 2D dataset. The analysis yielded a p-value of 0.006, an observed effect size of 0.531, and a statistical power of 0.817, supporting both statistical significance and practical relevance. In contrast, the post-exercise comparison between the datasets showed no statistically significant or practically meaningful difference, suggesting that any true effect is either absent or undetectable given the current sample size.

9.2.3 Limited benefit of combining pre- and post-exercise features



dataset	AUROC	95% CI	accuracy	95% CI	LR+	95% CI
2D	0.83	[0.68, 0.97]	0.74	[0.58, 0.87]	4.86	[1.73, 'n/a']
sweep	0.85	[0.71, 0.98]	0.68	[0.52, 0.84]	2.83	[0.9, 'n/a']
3D	0.91	[0.81, 1.0]	0.77	[0.61, 0.9]	5.46	[1.94, 'n/a']
TVR	0.9	[0.79, 1.0]	0.84	[0.71, 0.97]	12.14	[3.56, 'n/a']

Figure 9.5: Peripheral artery disease (PAD) classification using a random forest on the static 2D, sweep, and 3D datasets when both pre- and post-exercise feature values were combined. Results show no relevant improvement compared to models trained solely on post-exercise features (Fig. 9.4). The label 'TVR' refers to refinement performed via temporal volume registration (Sec. 5.1.2.4). LR^+ 'n/a' corresponds to 100% specificity (Eq. 2.33).

Combining pre- and post-exercise features did not substantially change classification performance. Extending the feature set to include both time points yielded results that were nearly identical to the best-performing configuration, namely post-exercise features alone, with only a slight improvement observed for the 3D dataset (Fig. 9.5).

Additional analysis of longitudinal pattern measurements before and after exercise, refined through temporal volume registration (Part II Sec. 5.1.2.4), likewise showed no notable gain in classification performance (Fig. 9.5).

9.2.4 Experimental demonstration of the importance of spatial matching

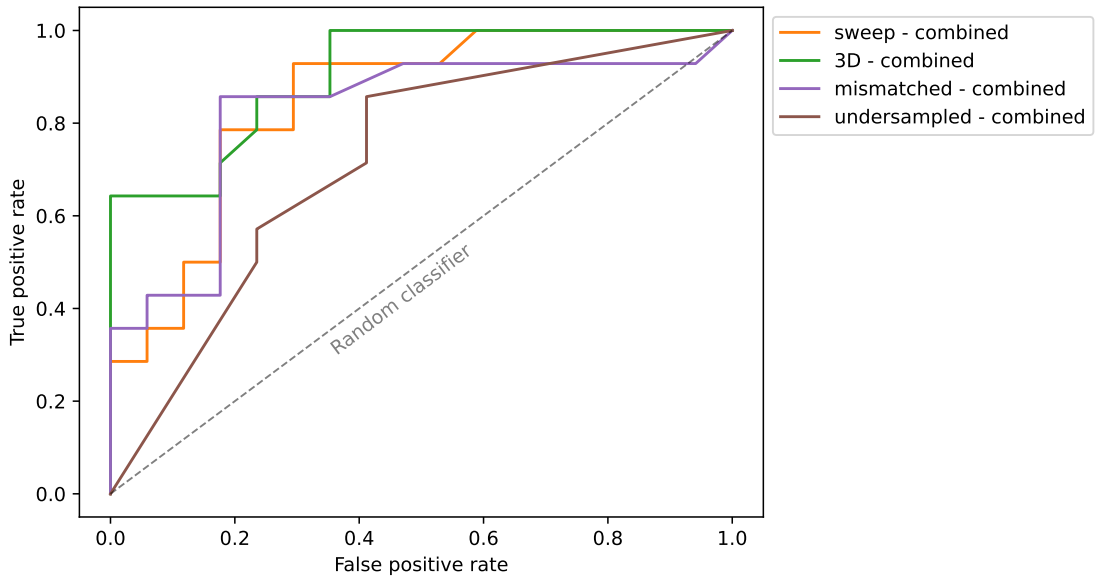
To qualitatively assess the influence of spatial matching, independent of variability introduced by using different scans for the static 2D and 3D datasets, a controlled toy experiment was conducted. The aim was to isolate the effect of spatial alignment from other confounding factors such as temporal signal variation, probe placement, and applied pressure.

For this experiment, only scans from the 3D dataset containing both pre- and post-exercise features were used. From these, four distinct sub-datasets were generated: (1) the semi-paired sweep dataset, (2) the fully paired 3D dataset, both as used in Section 9.2.3, (3) an undersampled dataset containing only the first, central, and last frame of each compounded 3D volume, and (4) a spatially mismatched dataset in which pre- and post-exercise frames were randomly selected from each volume, with a minimum enforced separation of 1 cm between the selected frames.

The results, shown in Fig. 9.6, demonstrate that the spatially mismatched and especially the undersampled datasets performed noticeably worse than the fully paired 3D dataset. However, the mismatched classification performance remained comparable to that of the original model trained on combined pre- and post-exercise 2D features (Fig. 9.5).

9.3 Discussion

This study presents the first in vivo application of optical pattern-based 3D PAI, demonstrating its feasibility and potential for clinical translation. A major outcome of this work is the successful PAD classification achieved using post-exercise static 2D and 3D PAI datasets by leveraging reperfusion and ischemic responses (H1). Similar classification results were achieved using pre-exercise 3D data, indicating the potential to detect muscular degeneration even before exercise (H2). In contrast, pre-exercise 2D classification performed at a level close to random guessing. This suggests that 3D



dataset	AUROC	95% CI	accuracy	95% CI	LR+	95% CI
sweep	0.85	[0.71, 0.98]	0.68	[0.52, 0.84]	2.83	[0.95, 'n/a']
3D	0.91	[0.81, 1.0]	0.77	[0.61, 0.9]	5.46	[1.92, 'n/a']
mismatched	0.82	[0.66, 0.99]	0.68	[0.52, 0.84]	2.83	[0.96, 'n/a']
undersampled	0.72	[0.54, 0.9]	0.68	[0.52, 0.84]	2.43	[0.99, 12.6]

Figure 9.6: Influence of spatial mismatch and undersampling on peripheral artery disease (PAD) classification. The experiment compares fully paired, semi-paired, undersampled, and spatially mismatched datasets to highlight the importance of spatial alignment. LR^+ 'n/a' corresponds to 100% specificity (Eq. 2.33).

imaging captures structural heterogeneities that conventional 2D imaging may miss, for example when measurements are acquired at suboptimal locations or do not adequately sample the affected tissue.

Pre- and post-exercise feature combinations further improved classification performance slightly in the 3D datasets, while no such improvement was observed in the static 2D datasets. A toy example qualitatively demonstrated that the advantage of pattern-based over conventional 2D imaging is likely driven by the spatial alignment achieved through the optical pattern registration.

Comparison between 3D and sweep datasets Interestingly, no substantial performance differences were observed between using the naïve image stack (sweep) and the optical pattern-based compounded 3D dataset for classification. There are likely three reasons for this. First, as validated in Part II Section 5.2.1.2, the 3D scans were already highly stable because the optical pattern was visible in the live MSOT view, providing visual guidance for the operator. Second, the sweep dataset may offer greater data diversity, as it contains frames captured before and after the optical pattern, which cannot be included in the spatially compounded 3D dataset but may still contain useful discriminative information. Third, the expressiveness of the compounded 3D dataset may be limited by interpolation, potentially yielding less realistic features from interpolated slices compared to original ones when used for PAD classification.

In line with this finding, applying temporal volume registration to the 3D datasets did not result in any noticeable improvement in classification performance. This suggests that minor spatial misalignments are negligible for PAD classification, consistent with prior 2D PAD studies where only approximate anatomical alignment, such as skin markings, was employed between pre- and post-exercise scans.

Detecting PAD pre-exercise 3D PAI is hypothesized to reveal muscular degeneration through spatial heterogeneity, enabling pre-exercise detection of PAD. This capability could be clinically valuable and potentially translate into direct patient benefit. A possible diagnostic pathway would involve performing a 3D scan at rest: if the result is positive for PAD, no further exercise-based testing would be necessary. If the result is negative, exercise could then be applied to assess reperfusion issues. This approach could reduce the time required by clinical staff and lessen the burden on patients by potentially avoiding the pain associated with exercise protocols, leading to direct patient benefit.

Feature combination Feature combination appears to be a promising strategy to improve PAD detectability, especially in the 3D datasets. However, this result need to be further validated due to the limited sample size and potential confirmation bias. It seems that feature combination preserves discriminative information from post-exercise measurements while incorporating baseline (pre-exercise) information.

Further limitations Several limitations of the present study should be noted. Each individual 3D measurement required approximately 30 seconds, meaning that the reconstructed volumes represent a temporal composite rather than an instantaneous

snapshot. In addition, the study design did not capture or standardize the interval between exercise completion and post-exercise measurement, which complicates interpretation of the physiological dynamics underlying the observed signal changes.

System-related factors may also have contributed to variability. Routine maintenance between the first and second measurement periods could have altered system characteristics, and six patients were measured with a N-shaped pattern that recovered only two translational DoF. Nevertheless, as established previously, the live visualization of the optical pattern provided sufficient guidance to obtain consistently high-quality pseudo-3D scans.

Finally, an important methodological limitation is that all analyses were conducted on the training set using LOO validation. The independent test set has not yet been evaluated, as the project is still ongoing and additional steps such as ROI detection and feature selection are in progress. As a result, the conclusions drawn here must be interpreted with caution, since their generalizability to unseen data has not yet been established. Confirmation on the independent test set will be essential to demonstrate the robustness and clinical relevance of the findings.

Confounding factors As indicated by the PCA analysis (Fig. 9.3 **C**), there appears to be a noticeable difference in sex-specific signal patterns. Therefore, I briefly detour into the analysis of potential confounding factors, which is especially, but not exclusively, relevant to avoid shortcut learning [230].

The work presented in this paragraph is based on work conducted by Marcel Knopp and Christoph J. Bender et al., in which I was involved as a co-author [231]. My contribution focused on data acquisition, as well as providing scientific advice and participating in the discussion of the experimental design and results.

Shortcut learning in this context refers to the risk that a machine learning model may not rely solely on disease-specific characteristics for classification, but instead may exploit confounding features such as sex or hardware-related noise patterns due to the changed membrane, especially when using imbalanced training sets (Fig. 9.7 **A**).

"Our study with 147 individuals demonstrates that [convolutional neural networks] CNNs can classify the sex from calf muscle [PAI] images, achieving an AUROC of 0.75. For PAD diagnosis, models trained on data with imbalanced sex-specific disease prevalence experienced significant performance drops (up to 0.21 AUROC) when applied to balanced test sets [(Fig. 9.7 **B** - orange vs. red tile)]. [...] [Therefore,] CNN-based models trained on [PAI] data may engage in shortcut learning by leveraging sex-related features, leading to biased and unreliable diagnostic predictions." [231]

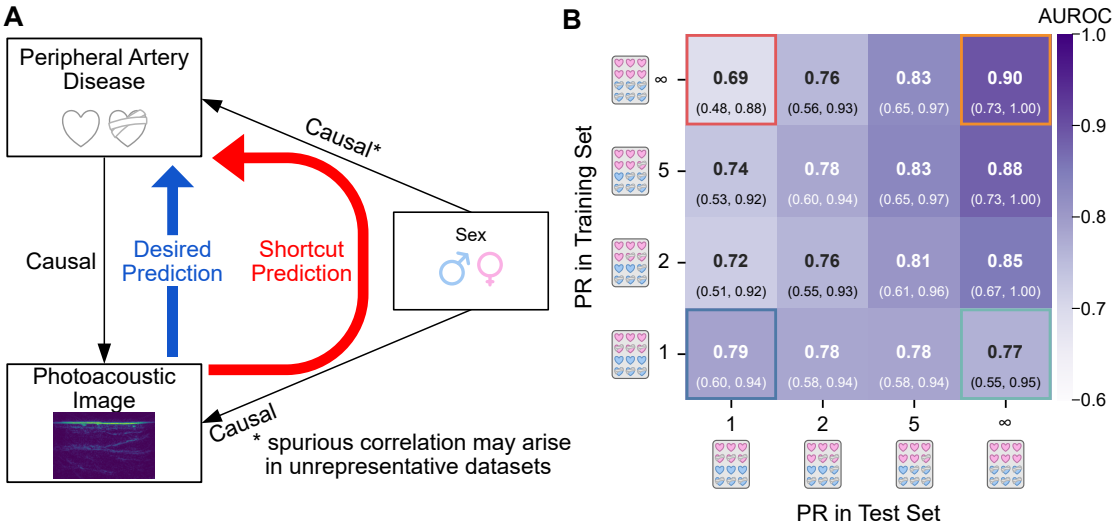


Figure 9.7: "**A** Illustration of shortcut learning in the context of peripheral artery disease (PAD)" [231]. **B** "Impact of sex-specific prevalence shifts from training set to test set. [PAD] classifiers trained on distributions with increasing sex-specific prevalence ratios (top to bottom PR =∞, 5, 2, 1) show significant performance drops and instability when tested on balanced data (PR = 1). Performance is measured with the mean area under the receiver operating characteristic curve (AUROC) shown in bold [...]. The model trained with PR =∞ experienced a 0.21 AUROC drop. Brackets [...] represent 95% confidence intervals" [231]. Figures adapted from [231] under CC BY 4.0 [120].

Similar effects were also observed when considering device-specific characteristics, such as membrane signals, laser energy variations, noise patterns, and device-specific artifacts [197]. These technical variances between devices can likewise lead to biased predictions and may introduce confounding factors in machine learning-based models.

An illustrative example of device-specific differences can be found in Part II Section 5.2.1, where the non-CE MSOT system (Fig. 5.10) exhibits pronounced ringing artifacts caused by electrical noise. In contrast, the CE MSOT system (Fig. 5.11) shows a substantially cleaner signal without such artifacts.

Related work and future directions Recent work by Träger et al. [112] demonstrated that muscle degeneration can be detected using single-wavelength PAI combined with ultrasound, further validated against histological examinations. This supports the hypothesis that pre-exercise 3D PAI is capable of detecting muscular degeneration.

Additionally, during the preparation of this thesis, a related study by our collaborators

was published [177]. They achieved near-perfect multi-threshold classification of PAD based on sO_2 using small, manually drawn ROIs in post-exercise measurements. However, their pre-exercise classification performance remained limited, with AUROC values ranging from approximately 0.62 to 0.74. These results reinforce the potential of 3D PAI to contribute to pre-exercise detection but also highlight the need for improved spatial localization of relevant regions. This suggests that future research should focus on implementing advanced segmentation methods, such as presented by Schillinger et al. [232], to improve ROI selection in 3D datasets. Automated segmentation would reduce the dependence on manual ROI placement, practically not feasible for 3D volumes, and may further enhance classification performance compared to the used fixed medium sized ROIs.

Conclusion This study further demonstrated that PAI is a promising modality for the diagnosis of PAD. Addressing the initial research question, no relevant performance difference was observed between 3D and 2D PAI in assessing reperfusion (H1) and, consequently, in PAD detection, while 3D PAI showed a clear advantage in detecting PAD through the identification of muscle damage (H2). However, no overall performance benefit of optical pattern-based compounded 3D volumes over naïve 2D image stacks was observed. Nevertheless, experimental findings (Sec. 9.2.4) suggest that the spatial and temporal context provided by 3D PAI may offer added value for PAD detection. This highlights the capability of 3D PAI as an initial diagnostic tool, which, by identifying muscular damage, may reduce the need for pain-inducing walking or heel-raise exercises, thereby lowering the physical burden on patients.

- Part IV: Closing -

10 Summary of contributions

PAI is a promising imaging modality that combines optical contrast with ultrasonic resolution, offering non-invasive and radiation-free access to functional and molecular information. Early studies have highlighted its potential in diverse clinical areas, ranging from oncology to vascular and inflammatory diseases. Nevertheless, adoption into routine care has not yet been achieved. This can be attributed to the fact that current implementations of PAI do not consider the relevant context around isolated images, which is crucial for a successful and reliable integration into existing clinical workflows.

10.1 Technical contributions

This thesis introduces the concept of *context-aware clinical PAI*, embedding context directly into acquisition and interpretation. Specifically, it considers (i) the spatial and temporal context required to draw meaningful conclusions about anatomy and physiology, (ii) the multimodal context that situates PAI alongside established modalities such as MRI and CT, complementing them with its unique functional contrast, and (iii) the biophysical context that links measured signals to both underlying physiology and the physics of image formation. By addressing these dimensions, context-aware PAI is reframed from a purely technical innovation into a workflow capable of generating reliable and interpretable clinical evidence.

The following section summarizes the technical contributions of this thesis, which span from optical pattern-based volumetric acquisition and multimodal fusion to digital-twin modeling for failure analysis, and describes how these advances were applied to selected clinical applications in alignment with the objectives defined at the outset (Part I Sec. 1.2). Relevant scientific dissemination is referenced where appropriate.

T1: Framework for volumetric PAI for clinical settings This work pioneers an optical pattern that encodes probe pose directly in each PA frame, enabling freehand 3D reconstruction and longitudinal alignment without external tracking. The framework is engineered for clinical practicality (minimal hardware, device-agnostic pipeline) and standardization (re-identifiable image planes and consistent acquisition quality). It turns orphaned 2D slices into reproducible 3D volumes and time-aligned series, a prerequisite for quantitative comparisons and longitudinal studies. By lowering setup complexity compared to motorized or tracking-based systems, it creates a viable path for bringing volumetric PAI into clinical research. The underlying concept of

the optical pattern has been published in a peer-reviewed journal as a first-author contribution [140], presented at SPIE photonics west 2021 [233], and the international conference on Information Processing in Computer-Assisted Interventions (IPCAI) 2021, where it received the *Audience Award for Best Innovation: Runner-Up* and *Best Paper: Honorary Mention*. The method has been successfully filed and granted as a European patent (2025) [184] and served as a foundational component in two successfully funded competitive research grants *SMART-3D* from the Bayerische Forschungsstiftung and *3D-PAD* from the ForTra Else Kröner-Fresenius-Stiftung (C2) [202].

T2: Marker-based multimodal 3D image registration This thesis presents the first marker-based approach that enables fusion of 3D PAI with established imaging modalities such as CT and MRI without relying on external tracking hardware. By extending the optical pattern framework developed in T1 with fiducials, PAI volumes could be co-registered to CT/MRI at millimeter accuracy, achieving performance comparable to optical tracking-based methods [161] while retaining a simple and device-independent setup. Such multimodal integration allows PAI to be complemented with anatomical detail beyond the resolution of PAI alone. This enables the transfer of planning information to the interventional setting and supports multimodal studies in which PAI contributes unique vascular and molecular contrast. These capabilities are key for clinician trust and for defining where PAI adds unique value. The method was presented as part of a poster presentation at the German Conference on Medical Image Computing (BVM) 2023 [234].

T3: Digital twin-based concept for failure analysis of clinical PAI For the first time, a clinical PAI digital-twin concept, designed to support analysis by disentangling physiological effects from image formation was presented. It enables identification of measurement limits, formulation of testable hypotheses for unexpected findings, and in principle also guidance in study design. In the head and neck RT study (C1), this approach helped explain counter-intuitive sO_2 trends as low BVF-induced estimation errors, separating biological effects from measurement limits. The twin provides a principled way to design studies, interpret outliers, and set expectations for PAI biomarkers. It could therefore become an essential step toward robust, explainable use of PAI leveraging biophysical context. This major contribution was published in the peer-reviewed journal Scientific Reports as a shared first-author contribution [93]. In this broader context, I also co-authored two peer-reviewed journal papers (1) contributing to the SIMPA toolkit [35], primarily to the processing components, as well

as (2) to a project focused on generating input masks for PA image synthesis from semantic segmentations [213].

10.2 Clinical applications

The broad clinical investigations in this thesis demonstrate, for the first time, how context-aware PAI can be deployed in real-world patient studies. Two complementary applications were chosen to test feasibility and reveal limitations: (C1) monitoring tissue response during radiotherapy in head and neck cancer, and (C2) detecting ischemia and chronic muscle damage in PAD. Together, they illustrate the potential and the boundaries of clinical PAI when augmented with spatial, temporal, and biophysical context.

C1: Photoacoustic imaging for monitoring radiotherapy treatment response in head and neck tumors Radiotherapy is central in treating head and neck tumors, yet early response assessment remains limited, as current evaluations rely on late anatomical changes. Because its effects first manifest functionally, PAI offers a non-invasive means to monitor such changes in real time. This thesis presents the first longitudinal clinical evidence that PAI can capture treatment-related changes during radiotherapy in humans. While the initial hypothesis of detecting reperfusion in malignant lymph nodes was not verified, the study uncovered tissue-level changes consistent with therapy side effects, such as declines in tHb and altered glandular water content. Crucially, the digital twin (T3) and LMM analyses explained counter-intuitive oxygenation trends by linking them to estimation uncertainty arising from low blood volume fractions and measurement variability, rather than underlying biological effects.

This work positions PAI in radiotherapy as an interpretable, context-aware method highlighting both, physiological responses and methodological limits. The results, published in Scientific Reports [93], set the stage for more robust, context-aware monitoring studies in oncology. Initial results on semantic segmentation in the head and neck region were co-presented by me as a poster at SPIE photonics west 2022 [224]. The methodological foundation for this work was established in the peer-reviewed journal publication *Semantic Segmentation of Multispectral Photoacoustic Images Using Deep Learning* [235], which I co-authored.

C2: Volumetric photoacoustics for detecting peripheral artery disease PAD is a prevalent condition that often presents with exercise-induced leg pain, yet current

diagnostic tools do not allow for reliable early diagnosis. Because ischemia and chronic muscle damage alter tissue oxygenation and structure, PAI offers a non-invasive means to assess these changes. Against this background, the second clinical contribution delivers the first in-human validation of the optical pattern framework (T1) in a vascular disease setting. Volumetric PAI successfully classified ischemia post-exercise (AUROC > 0.8), confirming feasibility. Importantly, while 2D imaging sufficed for post-exercise ischemia detection, only 3D imaging identified patterns consistent with chronic muscle damage in pre-exercise data, where 2D performance was near random. These findings establish a use-case boundary: 3D PAI is most impactful in scenarios where volumetric sampling and anatomical localization are essential, such as characterizing damage, while well-executed 2D imaging may suffice for acute ischemia screening. Initial pilot results were presented as a poster at SPIE photonics west 2023, where I received the 3rd prize in the *BiOS'23 3-Minute Poster Prize* competition [229]. Additionally, I co-authored related investigations in classification biases arising from sex [231] and hardware variability [197], highlighting the risk of shortcut learning in PAI.

10.3 Complementary contributions

In addition to the core work of this thesis, I contributed to several projects as co-author, reinforcing the concept of context-aware PAI. One line of work focused on quantitative oxygenation estimation, where Jan-Hinrich Nölke et al. introduced conditional invertible neural networks (cINNs) to replace single-value estimates with posterior distributions [50]. My contributions included methodological discussions and technical exchange. This project has the potential to add biophysical context by explicitly addressing uncertainty in sO_2 estimation. The importance of sO_2 estimation is underlined by another long-standing collaboration which began during my Bachelor's [36] and developed into a series of studies on spreading depolarizations in the gyrencephalic brain [37, 236, 237, 238]. Together with Thomas Kirchner and Edgar Santos et al., we demonstrated that PAI can capture rapid, transient sO_2 changes in the brain, providing temporal context for the study of acute pathophysiological events. More recently, I co-supervised two Master's theses that explored different aspects of context-aware PAI. Patricia Vieten investigated deep learning for semantic segmentation in PAI [223], a key step toward more realistic digital twin modeling. Sebastian Köpp applied the optical pattern in a large-animal model of Duchenne muscular dystrophy over a period of three months [239]. Within this project, I was deeply involved in grant

acquisition, planning, executing, and analyzing the experiments. This project tested the feasibility of longitudinal volumetric imaging in a preclinical setting and validated the robustness of the optical pattern under demanding translational conditions. Taken together, these complementary contributions illustrate how my collaborative work expanded context-aware PAI, while strengthening the methodological foundation for future clinical translation.

11 General discussion and outlook

This thesis set out to establish the concept of context-aware clinical PAI, embedding spatial and temporal, multimodal, and biophysical context into image acquisition and interpretation. In the following, these three dimensions are critically reflected with respect to their technical implementation, translational challenges, and clinical relevance. Building on these reflections, the chapter concludes by outlining what context-aware PAI now means in practice, what impact this work has achieved, and which directions appear most promising for future research.

11.1 Critical reflections

The following discussion reflects on the three dimensions of context addressed in this work (Fig. 1.1), considering their practical implementation as well as their strengths and limitations.

Spatial and temporal context This thesis has delivered a fully functional demonstrator of an optical pattern-based framework, spanning hardware and software, for generating 3D PAI data in clinical research. By embedding probe pose directly into the acquisition, the system adds spatial and temporal context and thereby enables reproducible, volume-based imaging.

Yet, translation into clinical routine will require not only regulatory clearance but also refinements in material properties (e.g., reducing acoustic damping) and tighter device integration. More fundamentally, the greatest barrier is not technical but clinical: no established standard-of-care use case yet demands 3D PAI. This chicken-and-egg dilemma, lacking adoption, because benefit is unproven, but unable to prove benefit without adoption remains unresolved. The true opportunity lies in identifying conditions where spatial compounding is indispensable, such as precise anatomical localization for monitoring dynamic physiological processes. In contrast, as the PAD study showed, high-quality 2D imaging can suffice, with the pattern framework serving primarily as a robust acquisition aid.

Multimodal context The optical pattern also creates an entry point for multimodal integration, embedding fiducials that enable registration of PAI with MRI and CT. Phantom experiments demonstrated sub-3 mm TRE, confirming technical feasibility and comparability with more complex tracking systems.

In practice, however, in vivo deviations due to tissue deformation reduce accuracy to a level suitable only for coarse alignment. Moreover, a clearly defined clinical need for such multimodal fusion is lacking. Discussions with clinicians highlighted that coordinating PAI with MRI sessions remains logistically heavy and offers little perceived benefit in routine workflows. Thus, while multimodal context has been convincingly demonstrated in principle, its translation will depend on scenarios where added functional information from PAI decisively complements established modalities. At present, it represents a promising but as yet unanchored capability.

Biophysical context The head and neck radiotherapy study exposed nearly every translational challenge in PAI: lack of standardized study designs, difficulties in semantic segmentation, and inherent image formation limits such as fluence variations and acoustic reflections. These issues culminated in puzzling results, like apparent post-therapy decrease in lymph-node sO_2. A retrospective digital twin analysis showed that low blood volume fraction could plausibly explain these anomalies as estimation errors rather than true physiology.

Nevertheless, modeling remains constrained by limited knowledge of how radiation alters tissue composition and by the persistent domain gap between simulated and real images. Current digital twins can highlight study limitations and guide planning but fall short of reliably predicting outcomes or fully explaining measured signals. Their role today is best seen as a critical lens for failure analysis and study design, not as a substitute for biological validation. The strength of biophysical context lies in exposing hidden assumptions, while its limitation is its immaturity as a predictive clinical tool.

11.2 Conclusion and future work

To conclude, this final section addresses the question: *What does "context-aware PAI" mean now, and what does it enable?* Foremost, PAI promises low-cost, non-invasive, radiation-free access to functional information, yet typical investigations still analyze isolated 2D slices without context.

This thesis tackles the major challenge of missing context by proposing a compact framework that (1) situates PA image slices in space and time, (2) complements gold-standard radiological imaging, and (3) analyzes when and why PA succeeds or fails, providing context across the clinical workflow. Additionally, this thesis presented the first longitudinal evidence in humans that PAI can monitor tissue response during radiotherapy, and a demonstration of the in-human feasibility of 3D PAI for ischemia detection in PAD.

These contributions advance the PA field by providing additional tools for the PAI toolbox. Collectively, *context-aware PAI* represents a workflow for reliable clinical PAI rather than merely a set of algorithms (Fig. 5.1). Specifically, it enables 3D multi-modal PAI with minimal hardware overhead and provides a systematic way to explain failure modes and measurement limits, guiding study design, data acquisition, and interpretation. Together, these contributions can help build clinical trust in PAI.

However, further investigations should focus on identifying compelling clinical use cases where added spatial, temporal, or multimodal context is indispensable, and on integrating context-aware PAI into the current clinical imaging landscape, enabling direct comparison with, and complementing, established modalities such as US, CT, and MRI with functional contrast. This, in turn, helps understanding where, when, and why PA measurements are beneficial, rather than whether they are.

Overall, this thesis establishes *context-aware clinical PAI* as a framework that embeds context into PAI workflows. This enhances both interpretability and clinical trust, and consequently marks a step toward clinical translation of PAI. With the identification of future high-impact indications, context-aware PAI could elevate the modality from exploratory imaging to an interpretable, clinically trusted complement to US, CT, and MRI.

'Everything hath an end,' said the Mayor,
'and that which we call a pudding hath two.'

Woodstock or The Cavalier
Walter Scott (1826)

Appendix

This appendix contains supplemental results, figures, and tables that support the findings presented in the main manuscript. The materials are organized in the same order as the main text and are grouped into three thematic chapters: (1) optical pattern, (2) monitoring of tissue response to RT, and (3) diagnosis of PAD.

A Optical pattern

Influence on ultrasound signal

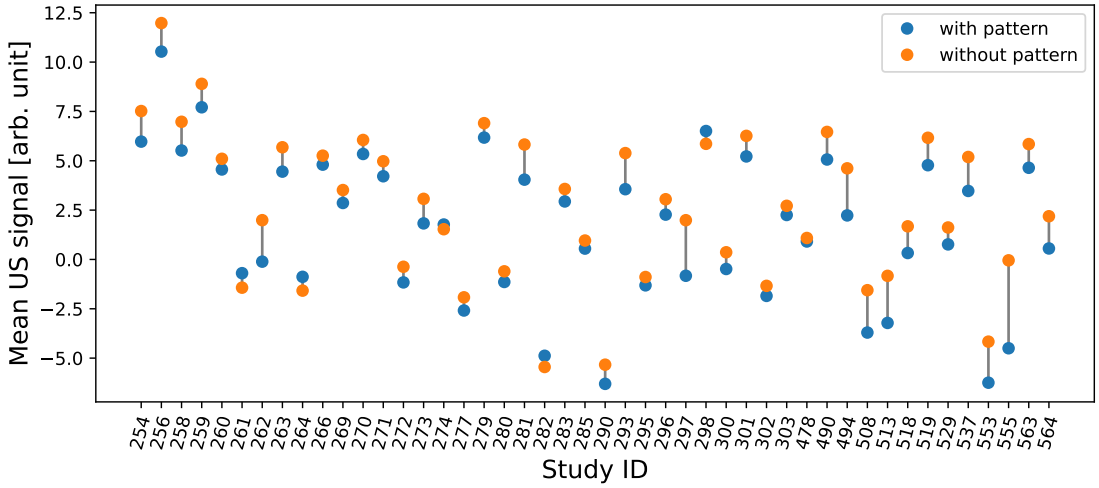


Figure A1: Comparison of mean ultrasound (US) values per individual, measured with (blue) and without (orange) the optical pattern. Each circle represents the mean US value from a single scan, calculated within a standardized region of interest (ROI) of 20 mm × 5 mm. An overall reduction in US signal is observed with the optical pattern.

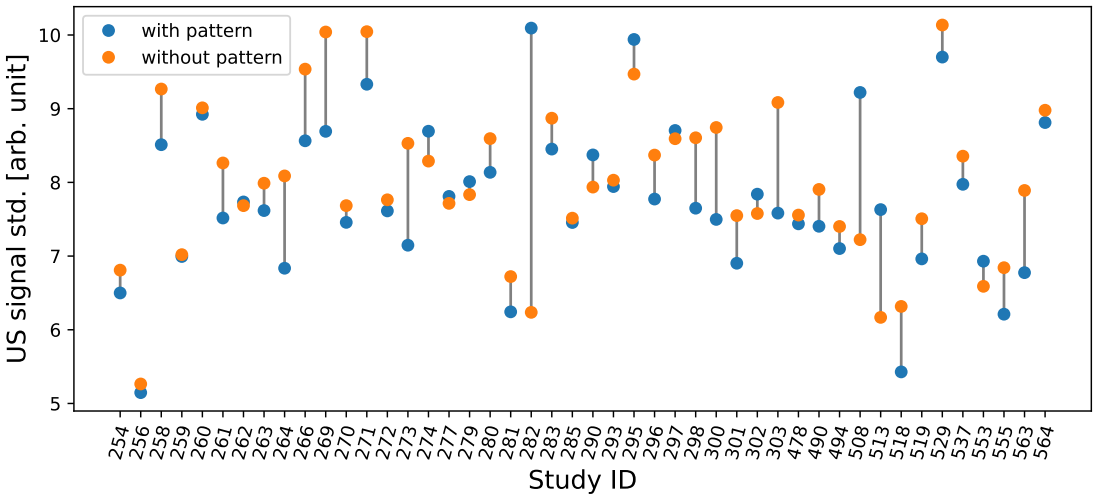


Figure A2: Comparison of the standard deviation of ultrasound (US) images per individual, measured with (blue) and without (orange) the optical pattern. Each circle represents the standard deviation of US values from a single scan, calculated within a standardized region of interest (ROI) of 20 mm × 5 mm. A decrease in US signal variability is observed in most scans with the optical pattern.

Influence flat surface constraint

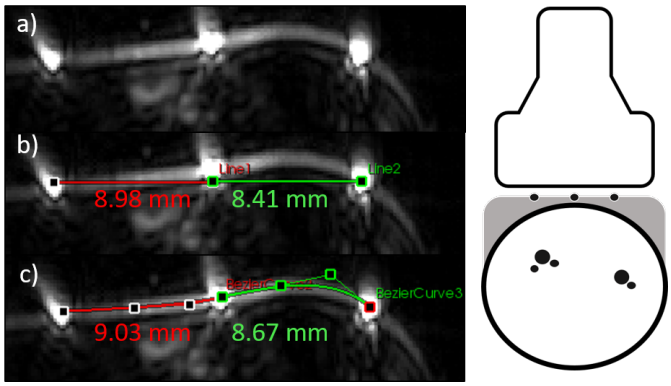


Figure A3: "Left: Analysis of an in-plane flat surface constraint violation. **a** Photoacoustic image of a healthy volunteer wrist showing the encoded pattern information. **b** Distances of the pattern absorption points assuming a flat surface. **c** Approximation of the real distances via a Bézier curve with two control points. Right: Potential hardware setup to overcome the flat surface constraint" [140]. Figure taken from [140] under CC BY 4.0 [120].

Background correction

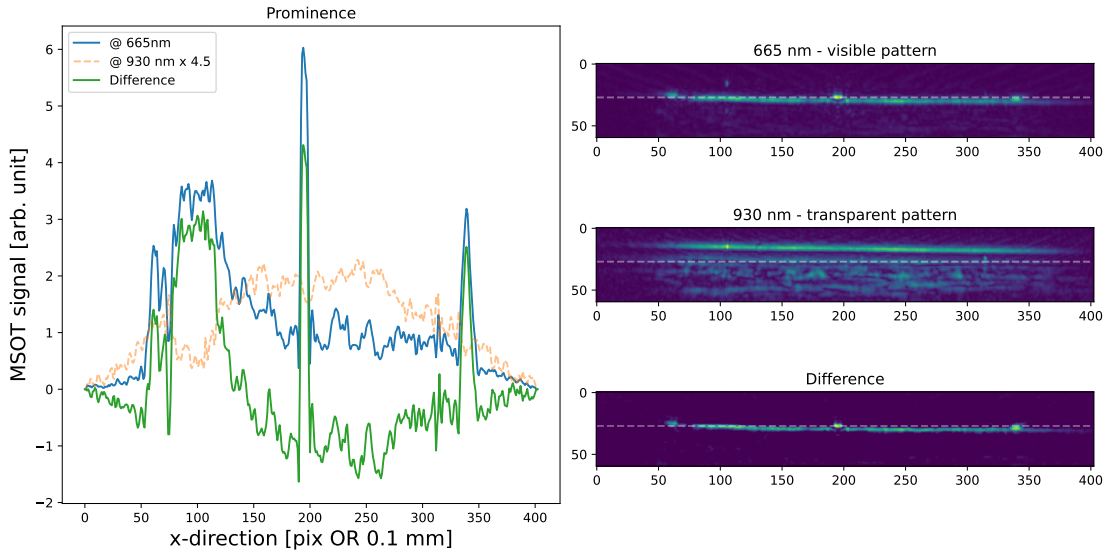


Figure A4: The suboptimal wavelength choice of 930 nm, where water and lipids exhibit higher absorption compared to 665 nm, reduces the effectiveness of background correction. **A** Line plot of photoacoustic (PA) signal intensity across the height of the pattern at 665 nm (blue) and 930 nm (dashed orange, scaled by a factor of 4.5), along with their difference (green), representing the result of the background correction. (**B**–**D**) Corresponding 2D PA images at 665 nm, 930 nm, and the corrected image, respectively.

Temporal volume registration and benchmarking

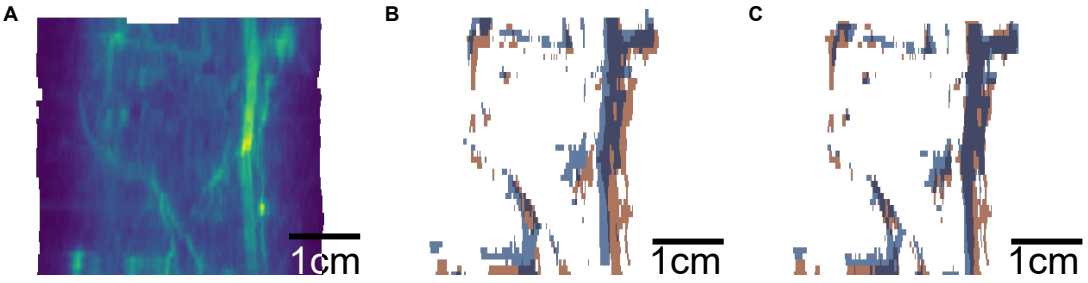


Figure A5: Rigid registration of temporally spaced in vivo 3D photoacoustic imaging (PAI) volumes of a human calf based on a 82nd signal quantile threshold segmentation. The Dice similarity coefficient increased from 0.55 to 0.61, indicating enhanced registration accuracy.

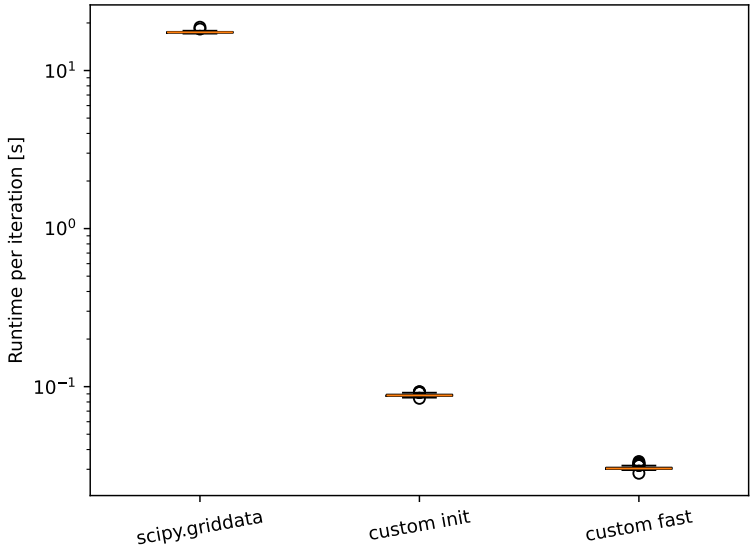


Figure A6: Qualitative benchmarking of 100 interpolation runs of a synthetic 3D "toy" volume with gradually varying pixel values. Each successive improvement in the custom interpolation implementation resulted in relevant speedup in average runtime (scipy.griddata 17.46 ± 0.22s; custom init 88.3 ± 1.5ms; custom fast 30.5 ± 0.7ms).

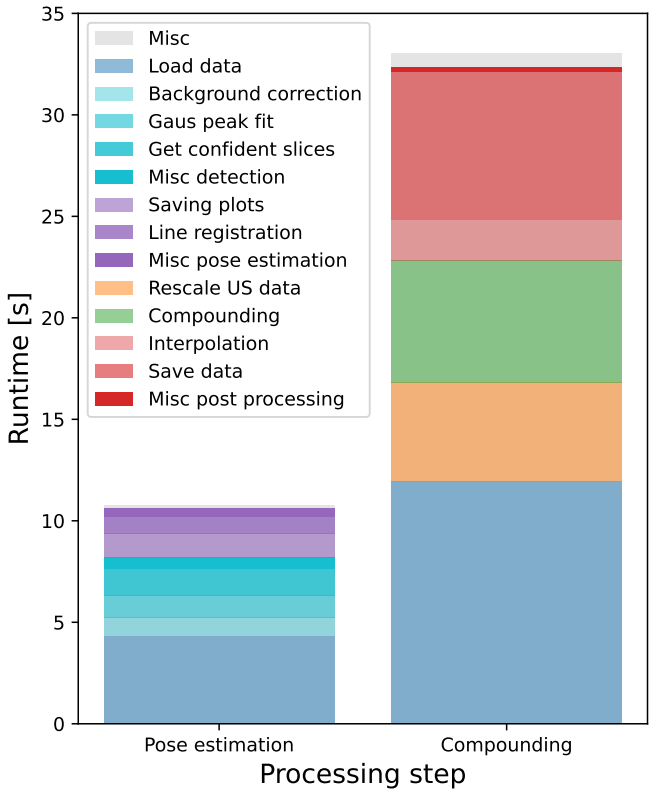


Figure A7: Benchmarking of the optical pattern-based framework, incorporating all processing steps, including ultrasound and unmixed data compounding. *Misc* refers to runtime elements that are not directly related to the core functionality of the framework, but instead cover aspects such as module interaction, initialization procedures, and other auxiliary processes.

Design

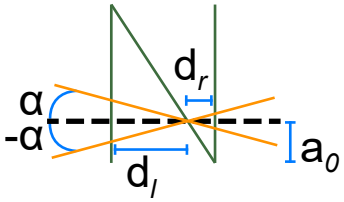


Figure A8: Geometric representation of an N-shaped optical pattern in the x–z view. The positions of the image slices (orange) can only be recovered in terms of their distance from the origin a_0 and the absolute angle α, but not the sign of α, since the left and right distances d_l and d_r are equal for both slices. This ambiguity in orientation makes accurate probe pose estimation infeasible with this design.

Multimodal image fusion

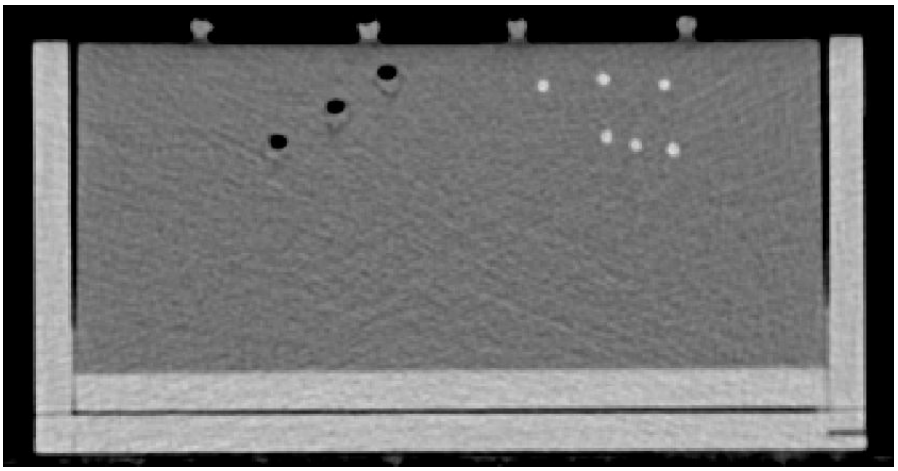


Figure A9: Tomographic view of a computed tomography (CT) image of the multimodal image fusion phantom. Visible are the pale grey enclosure, the darker grey agar background, and four fusion fiducials positioned on top of the volume. On the left, three target markers are placed at different depths, where air inclusions appear as distinct black dots due to the marker fluid having similar density to the agar. On the right, two layers of N-shaped wires filled with nigrosin solution are detectable as bright white spots.

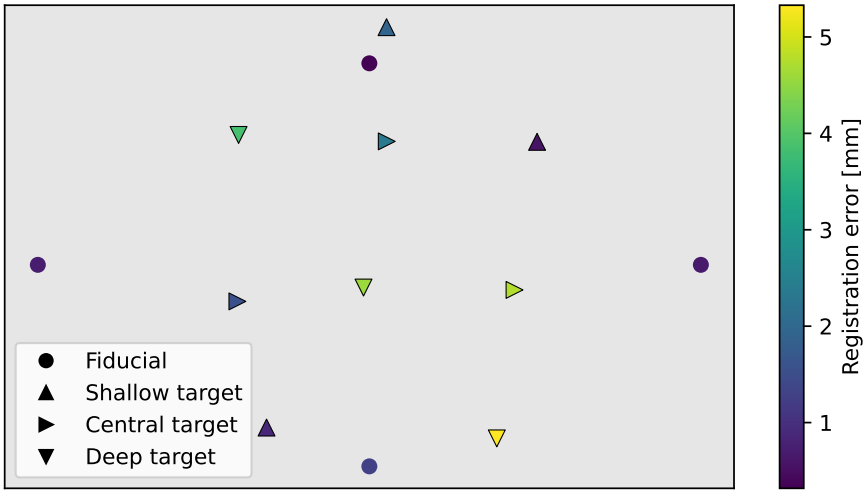


Figure A10: Post-registration x-z projection of all nine target markers (triangles) and four fiducial markers (●) for computed tomography (CT) photoacoustic imaging (PAI) image fusion. Triangle orientation encodes assigned depth levels: shallow (↑), central (→), and deep (↓), corresponding to E1-E3 depth categories (Fig. 6.3). Marker color indicates the local registration error.

B Monitoring of tissue response to radiotherapy

The following appendix chapter is taken from my first-author publication [93]. All content, including figures, captions, and descriptions, remains unchanged from the original work.

LN	700 nm		730 nm		760 nm		800 nm		850 nm	
	min	max	min	max	min	max	min	max	min	max
μ_a	0.069	1.410	0.0590	1.052	0.083	1.510	0.052	1.180	0.074	1.380
μ_s	6.750	235.250	6.5007	226.250	6.260	217.930	5.970	207.780	5.640	196.390
g	0.994	0.999	0.9940	0.999	0.994	0.999	0.994	0.999	0.994	0.999

Fat	700 nm		730 nm		760 nm		800 nm		850 nm	
	min	max	min	max	min	max	min	max	min	max
μ_a	0.060	0.061	0.053	0.053	0.076	0.078	0.057	0.06	0.078	0.078
μ_s	58.410	60.630	57.035	59.440	55.750	58.330	54.190	56.96	52.410	55.390
g	0.968		0.968		0.968		0.968		0.968	

ST	700 nm		730 nm		760 nm		800 nm		850 nm	
	min	max	min	max	min	max	min	max	min	max
μ_a	0.060	0.255	0.052	0.198	0.076	0.284	0.057	0.224	0.078	0.270
μ_s	43.120	72.340	41.330	69.450	39.709	66.810	37.760	63.620	35.640	60.091
g	0.897	0.901	0.897	0.901	0.897	0.901	0.897	0.901	0.897	0.901

Table B1: "Optical properties absorption coefficient (μ_a [cm^{-1}]), scattering coefficient (μ_s [cm^{-1}]), and anisotropy (g) for lymph node (LN), fat, and soft tissue (ST) used for the digital twin model." [93] Table from [93] under CC BY 4.0 [120].

Epidermis	700	730	760	800	850
μ_a [cm^{-1}]	3.0	2.0	2.0	1.0	1.0
μ_s [cm^{-1}]	52.0	49.0	47.0	44.0	41.0
g	0.804	0.813	0.823	0.835	0.849

Heavy water	700	730	760	800	850
μ_a [cm^{-1}]	0.001	0.001	0.001	0.001	0.001
μ_s [cm^{-1}]	1.0	1.0	1.0	1.0	1.0
g	1.0	1.0	1.0	1.0	1.0

Mediprene	700	730	760	800	850
μ_a [cm^{-1}]	0.017	0.017	0.017	0.017	0.017
μ_s [cm^{-1}]	0.147	0.147	0.147	0.147	0.147
g	0.9	0.9	0.9	0.9	0.9

US gel	700	730	760	800	850
μ_a [cm^{-1}]	0.006	0.018	0.025	0.02	0.043
μ_s [cm^{-1}]	1.0	1.0	1.0	1.0	1.0
g	1.0	1.0	1.0	1.0	1.0

Table B2: "Optical properties absorption coefficient (μ_a), scattering coefficient (μ_s), and anisotropy (g) for the epidermis, heavy water (couplant), and ultrasound (US) gel used for the digital twin model." [93] Table taken from [93] under CC BY 4.0 [120].

	Density [kg/m^3]	SOS [m/s]	Ac. attenuation [dB/cm/MHz]
Epidermis	1109.0	1624.0	0.35
Fat (min)	975.5	1473.1	0.084
Fat (max)	982.2	1480.6	0.11
US gel	1000.0	1482.30	0.003
Heavy water	1107.0	1540.0	0.003
Mediprene	890.0	1583.0	0.277
Soft tissue (min)	1000.5	1501.2	0.063
Soft tissue (max)	1002.5	1502.7	0.069
Lymph node (min)	1032.8	1584.7	2.12
Lymph node (min)	1037.4	1587.1	2.83

Table B3: "Acoustic properties density, speed of sound (SOS), and alpha coefficient/acoustic attenuation used for the digital twin model. US: ultrasound" [93] Table taken from [93] under CC BY 4.0 [120].

Lymph node ID	Location	Level	T	N	M	AJCC V8	Age	Sex
1	oropharynx	IIA	2	2	0	II	59	m
4/5	oropharynx	IIA/III	2	2	0	II	57	m
7	hypopharynx	III	3	3	0	IVB	73	m
8/10	oropharynx	IIA/III	4	3	1	IVB	66	m
11	oropharynx	V	3	2	0	II	66	m
15	oropharynx	IIB	4	2	0	III	64	m
16/19	hypopharynx	IIA/III	4	2	0	IVB	54	f
20	oral cavity	IIA	1	1	0	III	59	m
22/23	larynx	IIA/III	3	1	0	III	69	m
25	oropharynx	IIA	4	1	0	III	69	m
28/29	oropharynx	III/III	3	2	0	II	66	m

Table B4: "Patient characteristics: The table gives an overview of the patient characteristics of the investigated tumors and the corresponding lymph nodes [...]. Histologically all tumors were classified as squamous cell carcinomas in the head and neck area. The locations of the primary tumors were heterogeneous with a dominance in the oropharyngeal space. The level describes the location of the lymph node metastases following the classification system of lymph node levels in head and neck cancers after Robbins et al. [240]. Tumor, lymph node, metastasis (TNM) classification and staging after the American Joint Committee on Cancer (AJCC) 8th edition [241] shows the extent of disease when patients were diagnosed and included in the study. The age at first diagnosis as well as the gender of the patient are also represented." [93] Table taken from [93] under CC BY 4.0 [120].

Node ID	Size pre	Depth pre	Size during	Depth during	Size post	Depth post
1	0.9	2.5	1.1	0.6	0.9	2.3
4	0.7	0.9	0.7	1.0	1.2	0.9
5	0.8	1.4	1.3	1.3	0.9	1.4
7	1.5	0.9	1.4	0.4	0.9	1.6
8	1.5	1.5	1.3	1.7	1.0	1.4
10	1.5	0.6	1.8	0.6	1.2	1.2
11	1.0	0.6	1.2	1.0	1.1	0.6
15	1.7	0.9	1.6	0.8	0.4	1.8
16	1.2	0.8	0.6	2.4	0.5	0.6
19	0.7	1.3	0.8	1.1	0.7	1.0
20	1.8	0.7	1.1	0.6	0.8	0.9
22	1.2	1.2	1.0	1.1	0.6	1.4
23	0.7	1.7	0.8	1.8	1.0	0.8
25	1.4	1.3	1.3	1.8	1.1	1.6
28	1.0	1.1	0.8	1.6	0.8	1.2
29	1.1	0.4	0.8	0.6	1.0	1.2

Table B5: "The table gives the lymph node size in cm^2 and depth in cm at each measurement before (pre), during (after the 9th fraction), and 90 days after the radiotherapy. The values were derived from the semantic segmentation masks." [93] Table taken from [93] under CC BY 4.0 [120].

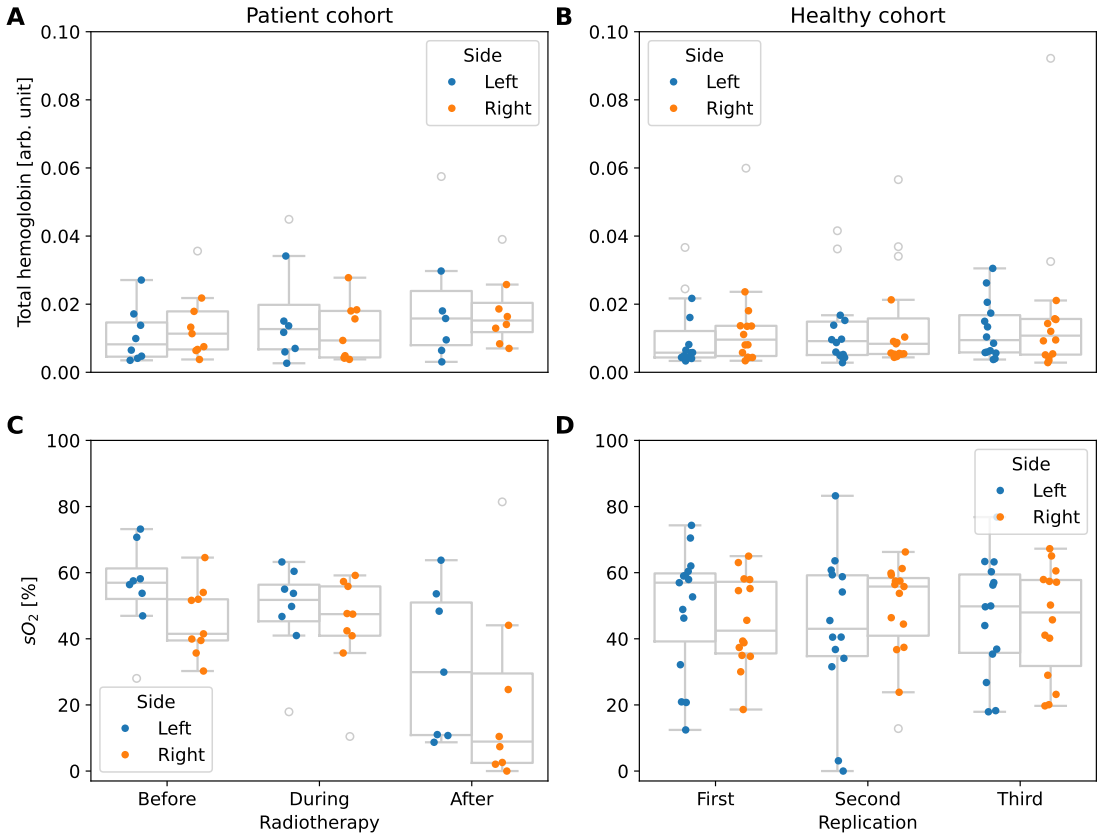


Figure B1: "Biomarker distributions over time for total hemoglobin (A/B), as well as oxygen saturation (sO_2) (C/D), are shown for malignant patient (A/C) and benign healthy volunteer (no radiotherapy - B/D) lymph nodes. The box shows the interquartile range (IQR) with the median as center line. The whiskers extend to points that lie within 1.5 * IQR of the 1st or 3rd quantile. All data points are plotted on top, and color-coded by measurement side (left or right neck). Outliers are shown without color." [93] Figure taken from [93] under CC BY 4.0 [120].

Time	Side	Depth	Counts	Smoking	BMI	WS	Subject-Lymph	Unexplained
32	3	1	-1	-3	-3	0	32	40

Table B6: "We illustrate the variance decomposition for the endpoint oxygenation [%] derived from the linear mixed model (LMM) analysis. The variability in oxygen saturation (sO_2) in malignant patient lymph nodes is mainly unexplained, due to inter-target variability (subject-lymph) and due to time/treatment. Location (side), depth, pixel in the region of interest (counts), smoking history, body mass index (BMI), and season of measurement (WS) do not play a relevant role."[93] Table taken from [93] under CC BY 4.0 [120].

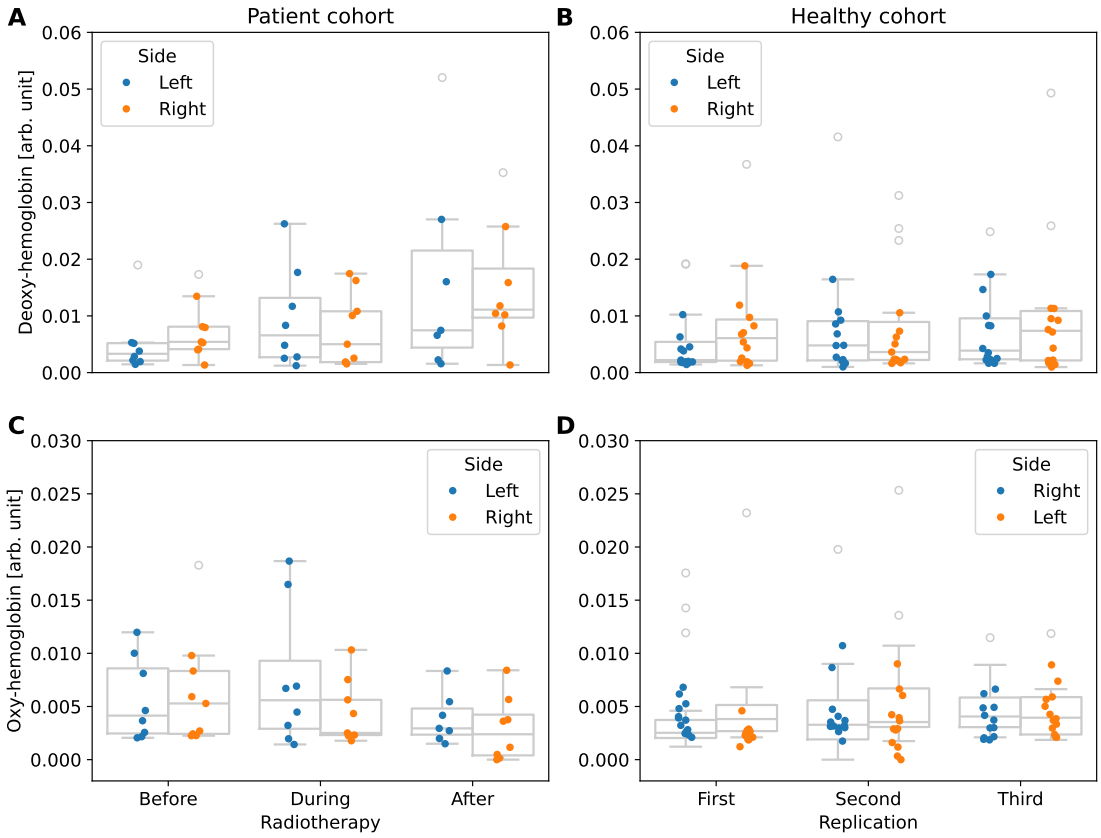


Figure B2: "Biomarker distributions over time for deoxy- (A/B), oxy- (C/D) hemoglobin, are shown for malignant patient (A/C) and benign healthy volunteer (no radiotherapy - B/D) lymph nodes. The box shows the interquartile range (IQR) with the median as center line. The whiskers extend to points that lie within 1.5 * IQR of the 1st or 3rd quantile. All data points are plotted on top, and color-coded by measurement side (left or right neck). Outliers are shown without color." [93] Figure taken from [93] under CC BY 4.0 [120].

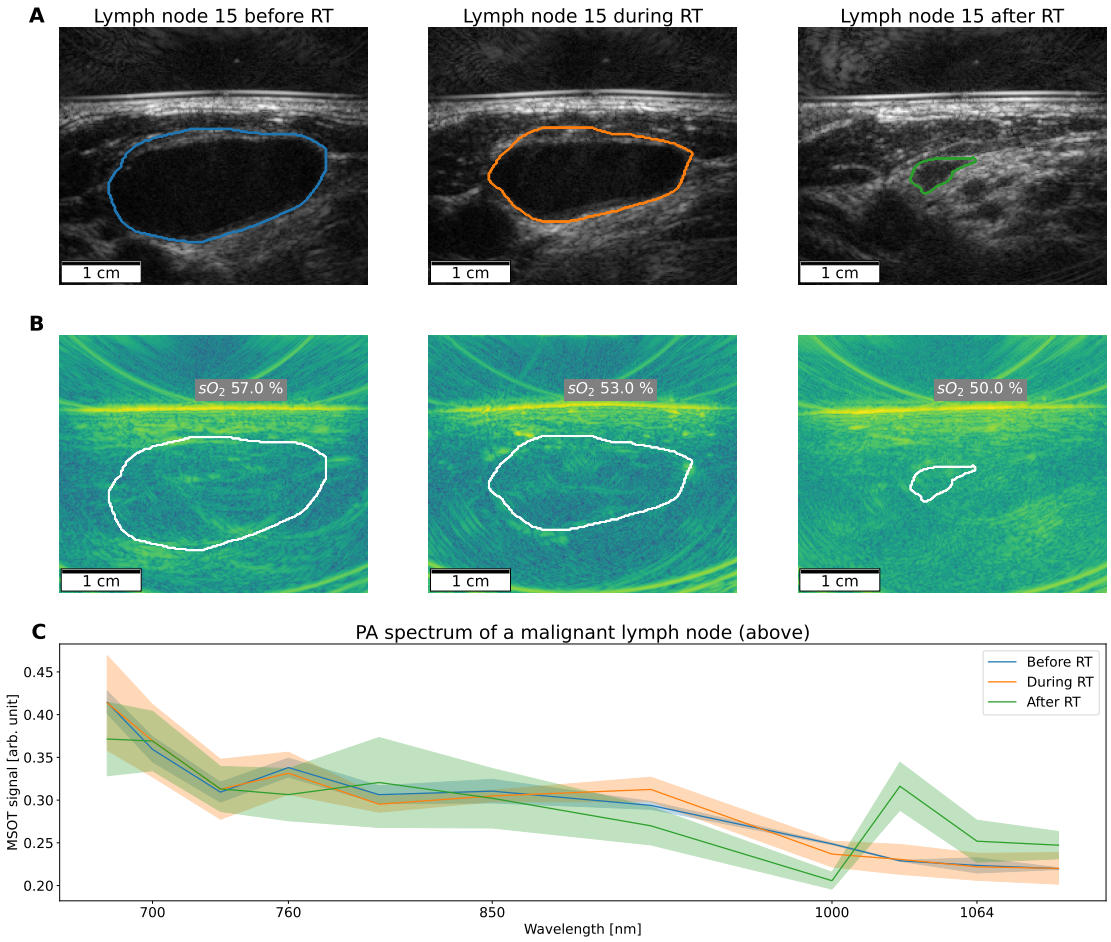


Figure B3: "Example lymph node (ID 15) over time with corresponding ultrasound (US) images (A), photoacoustic (PA) images (B), and spectral information with standard deviation (C) derived from the region of interest (semantic segmentation of node as contour). Additionally, the derived oxygen saturation (sO_2) over the course of the radiotherapy (RT) is displayed in the PA image. The shrinking of the lymph node metastasis within three months can be attributed to a positive therapeutic response." [93] Figure taken from [93] under CC BY 4.0 [120].

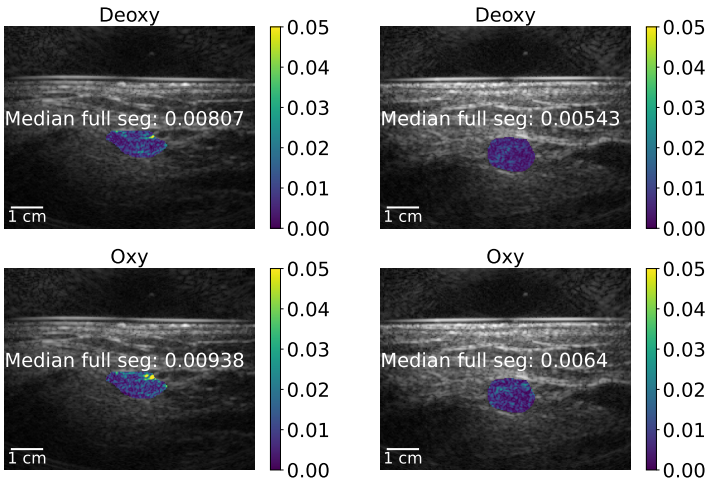


Figure B4: Patient lymph node with tumor progression before (left) and during (right) therapy, showing deoxy- and oxyhemoglobin signals (rows). Oxygen saturation (sO_2) levels remain constant at 54% pre- and during therapy. Limited tissue activity is observed, and overall image quality is suboptimal. In each figure, the median value of the full photoacoustic (PA) signal within the annotated lymph node (colored) is shown, overlaid on the grayscale ultrasound (US) background image.

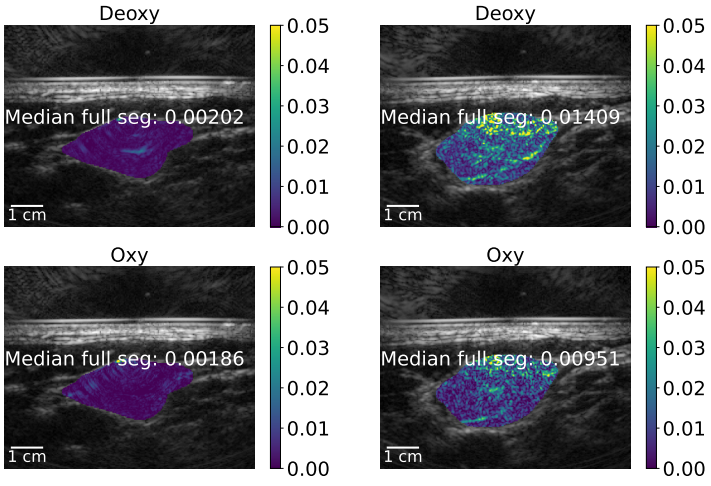


Figure B5: Patient with successful therapy before (left) and during (right) treatment, showing deoxy- and oxyhemoglobin signals (rows). High tissue activity is visible, but a drop in oxygen saturation (sO_2) is observed (48% before and 40% during therapy). In each figure, the median value of the full photoacoustic (PA) signal within the annotated lymph node (colored) is shown, overlaid on the grayscale ultrasound (US) background image.

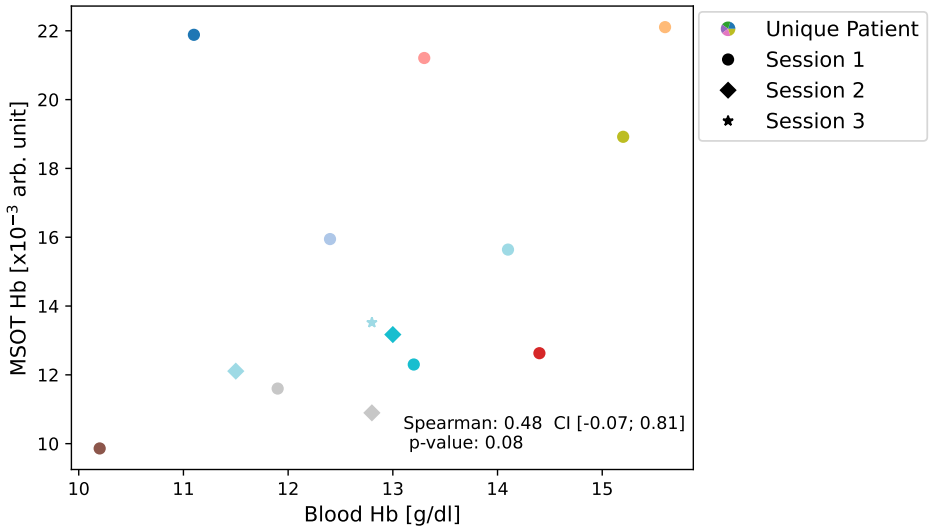


Figure B6: Correlation between total hemoglobin (tHb) levels derived from the photoacoustic (MSOT) signal (y-axis) and clinical blood samples (x-axis), color-coded by patient (11 individuals) and marked according to the time point relative to radiotherapy: before (circle), during (diamond), and after (star) treatment. The MSOT-derived tHb value represents the mean tHb signal across all measured organs for each patient at the corresponding time point. A Spearman correlation coefficient of 0.48 (95% CI: [–0.07, 0.81]; p = 0.08) indicates a moderate correlation. Appendix Figure B7 shows the results for only the first timepoint, accounting for the hierarchical structure of the data.

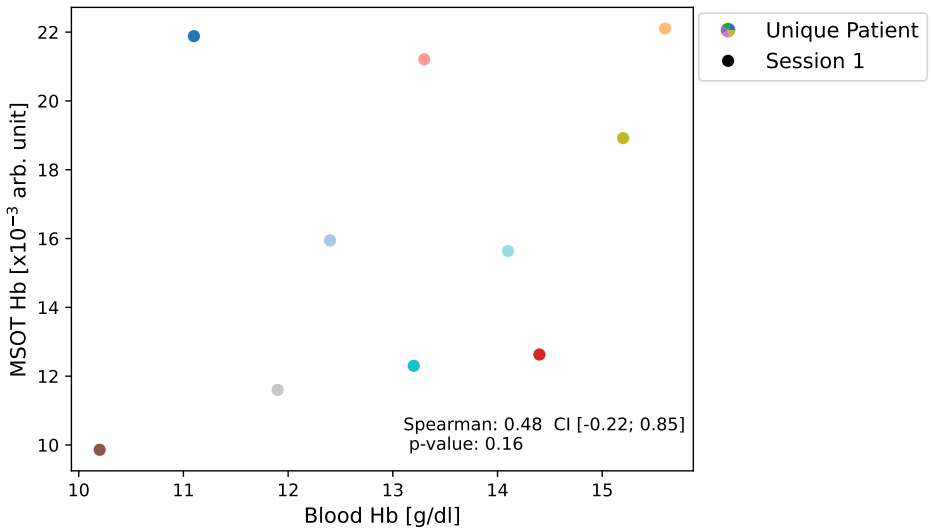


Figure B7: Correlation between total hemoglobin (tHb) levels derived from the photoacoustic (MSOT) signal (y-axis) and clinical blood samples (x-axis), color-coded by patient (11 individuals) before treatment. The MSOT-derived tHb value represents the mean tHb signal across all measured organs for each patient. A Spearman correlation coefficient of 0.48 (95% CI: [–0.22, 0.85]; p = 0.16) indicates a moderate correlation.

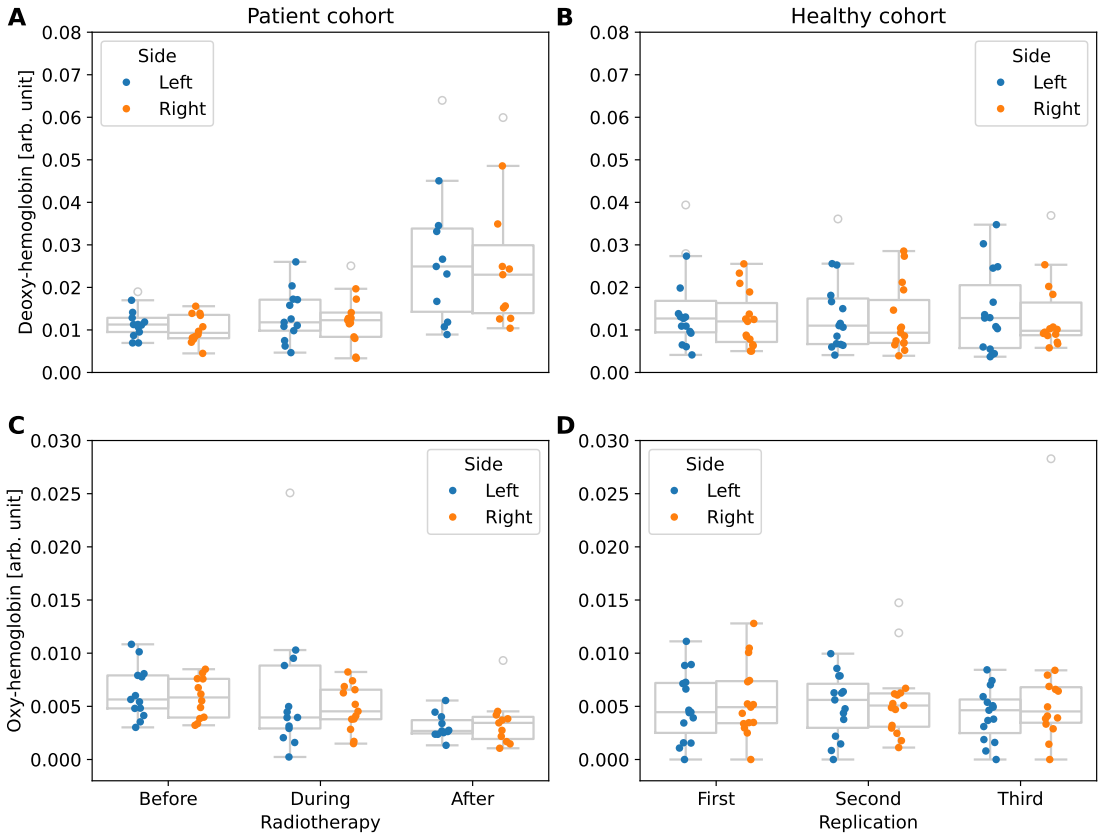


Figure B8: "Strong increase in total hemoglobin within the sternocleidomastoid during radiotherapy (RT) is mainly caused by deoxyhemoglobin increase (A) and decrease in oxyhemoglobin (C) for patients undergoing RT treatment. For comparison, the healthy volunteer data (no RT) are displayed (B/D). The box shows the interquartile range (IQR) with the median as center line. The whiskers extend to points that lie within 1.5 * IQR of the 1st or 3rd quantile. All data points are plotted on top, and color-coded by measurement side (left or right neck). Outliers are shown without color." [93] Figure taken from [93] under CC BY 4.0 [120].

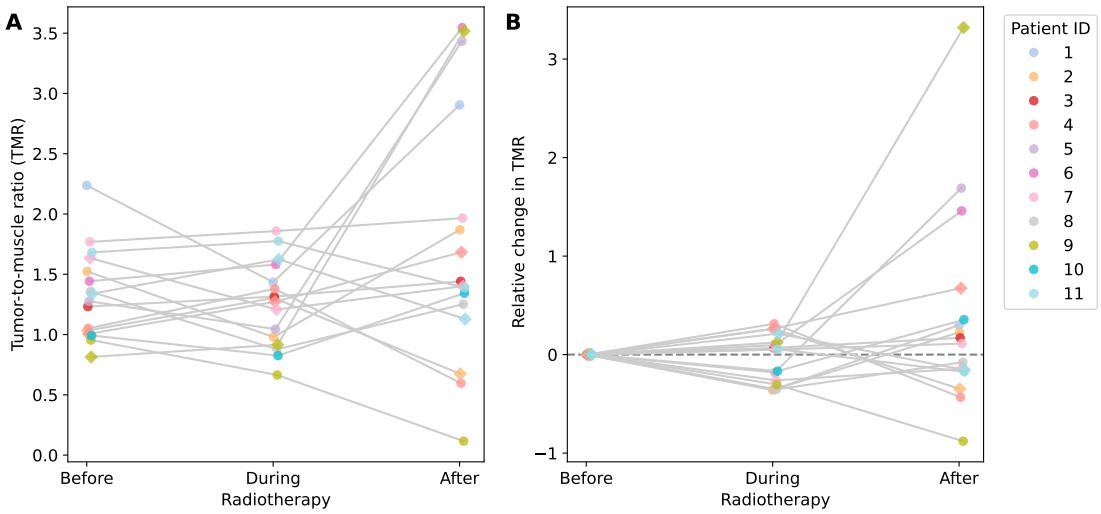


Figure B9: "Tumor-to-muscle ratio (TMR) of oxygen saturation (sO_2) increases in suspect lymph nodes throughout radiotherapy (RT). This is shown for the overall TMR values (A) and relative to the first session (B), with each node color-coded by patient. For some patients, multiple nodes were measured, as indicated by their shape (1st node: circle; 2nd node: diamond shape)." [93] Figure taken from [93] under CC BY 4.0 [120].

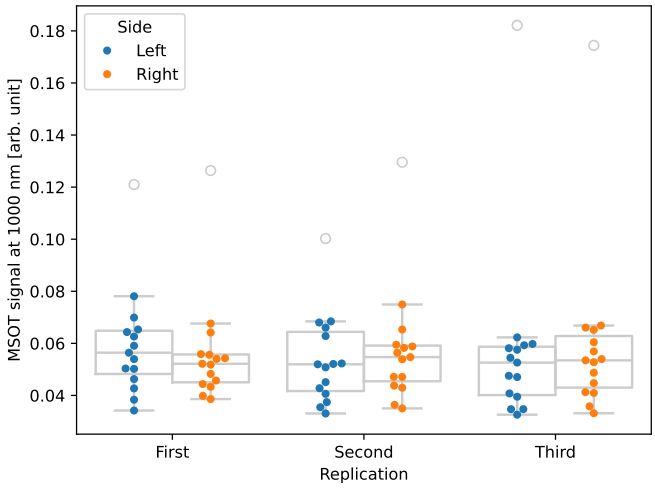


Figure B10: "Photoacoustic signal of the submandibular gland at 1000 nm (near the water absorption peak) of healthy volunteers for all three repetitions. The box shows the interquartile range (IQR) with the median as center line. The whiskers extend to points that lie within 1.5 * IQR of the 1st or 3rd quantile. All data points are plotted on top, and color-coded by measurement side (left or right neck). Outliers are shown without color." [93] Figure taken from [93] under CC BY 4.0 [120].

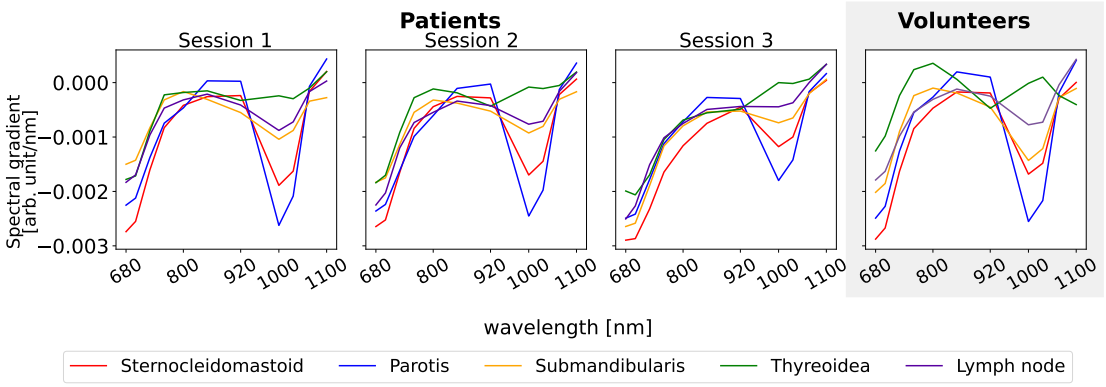


Figure B11: Mean spectral gradient (Eq. 8.2) for five target anatomical structures, sternocleidomastoid muscle, parotid gland, submandibular gland, thyroid gland, and lymph nodes with light purple representing benign and dark purple representing suspect malignant nodes. From left to right, spectra are shown for three imaging sessions (before, after the 9th fraction, and 90 days after radiotherapy) followed by the spectral gradients from healthy volunteers.

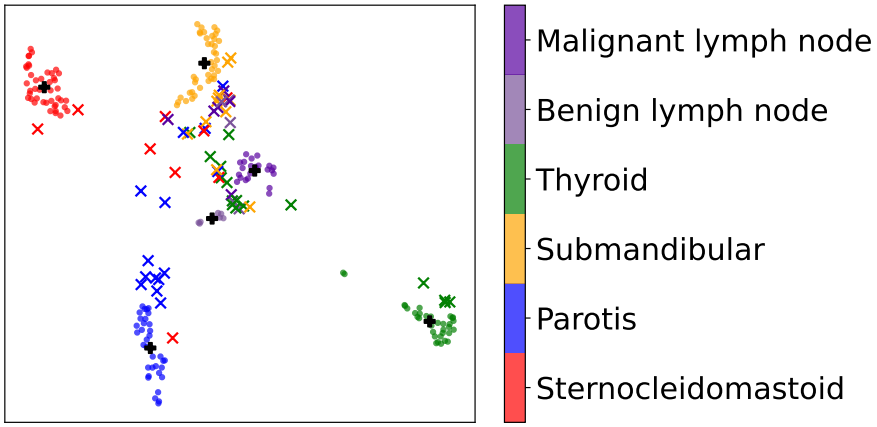


Figure B12: Supervised uniform manifold approximation and projection for dimension reduction (UMAP) clustering of spectral data from various head and neck organs (color-coded circles). Each point represents the spectral signature of tissue sampled from a specific anatomical structure, embedded into a 2D space using UMAP, with supervision provided by labeled tissue types. The clear separation between most classes indicates distinct spectral characteristics across the organs, supporting the feasibility of organ-level spectral classification. Cluster centroids, computed using the KMeans algorithm, are shown in black (+). Notably, the clustering does not generalize to the 25% unseen test data (x) and also fails to yield meaningful results in a fully unsupervised setting.

C Diagnosis of peripheral artery disease

Step: Selecting multi-class counting metrics (S2)	
Which cutoff method on predicted class scores fits your problem more?	Optimization-based
Are class confusions of unequal severity in your study?	No
Is there unequal interest across classes in your research?	No
Do the provided class prevalences reflect the population of interest?	Yes
Is there a class imbalance in your research?	No
Selected Metric:	Accuracy
Step: Selecting per-class counting metric (S3)	
Which per-class metric better suits your problem? Options: - LR^+; sensitivity: F_β Score	LR^+
Selected Metric:	LR^+ (per class)
Step: Selecting multi-threshold metric (S4)	
Are predicted class probabilities available in your research?	Yes
Which multi-threshold metric better suits your problem? Options: - AUROC; average precision	AUROC
Selected Metric:	AUROC
Step: Selecting calibration metric (S5)	
Is calibration assessment requested in your research?	No
Selected Metric:	None

Table C1: Output responses from the Metrics Reloaded toolkit [68] for the *image level classification* problem category. Abbreviations: positive likelihood ratio (LR^+); area under the receiver operating characteristic curve (AUROC).

Parameter	Value	Parameter	Value
`n_estimators`	100	`random_state`	66
`max_depth`	None	`min_samples_split`	2
`min_samples_leaf`	1	`min_weight_fraction_leaf`	0.0
`max_features`	1.0	`max_leaf_nodes`	None
`min_impurity_decrease`	0.0	`bootstrap`	True
`oob_score`	False	`n_jobs`	None
`verbose`	0	`warm_start`	False
`ccp_alpha`	0.0	`max_samples`	None
`criterion`	'gini'	`monotonic_cst`	None

Table C2: Parameters used for `sklearn.ensemble.RandomForestClassifier` in peripheral artery disease (PAD) classification.

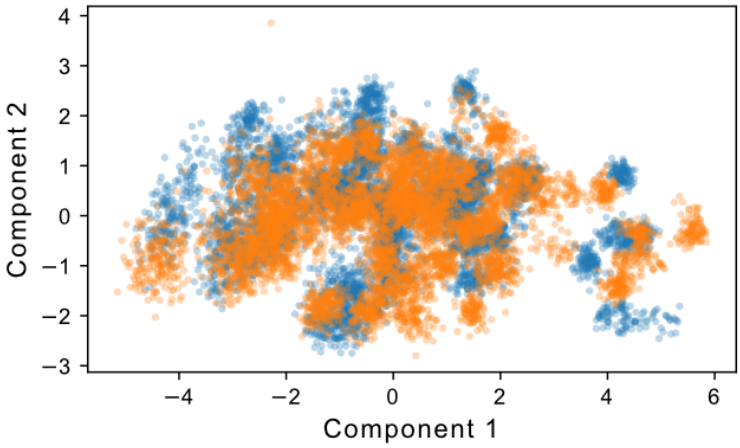


Figure C1: Principal component analysis (PCA) (`sklearn.decomposition`) of the training set from the peripheral artery disease (PAD) follow-up study. Each dot represents the derived principal components, calculated from the mean value within a 20 mm x 5 mm cropped region of interest of one frame of one scan across all functional wavelengths (730, 760, 800, 850, 930, 1030 nm). Color coding indicates exercise condition: blue before and orange after exercise (heel raises).

Feature	Dataset	Exercise	AUROC	CI 95%
800 nm	sweep	post	0.64	[0.50 – 0.78]
800 nm	3D	post	0.62	[0.47 – 0.77]
850 nm	2D	pre	0.62	[0.46 – 0.77]
850 nm	2D	post	0.76	[0.63 – 0.89]
850 nm	sweep	pre	0.62	[0.47 – 0.77]
850 nm	sweep	post	0.70	[0.56 – 0.84]
850 nm	3D	pre	0.64	[0.48 – 0.79]
850 nm	3D	post	0.74	[0.60 – 0.88]
sO_2	2D	post	0.67	[0.52 – 0.81]
sO_2	sweep	post	0.62	[0.47 – 0.77]
sO_2	3D	post	0.64	[0.49 – 0.80]
HbO_2	2D	post	0.76	[0.63 – 0.89]
HbO_2	sweep	pre	0.68	[0.55 – 0.82]
HbO_2	sweep	post	0.76	[0.62 – 0.89]
HbO_2	3D	pre	0.70	[0.57 – 0.84]
HbO_2	3D	post	0.74	[0.61 – 0.88]

Table C3: Peripheral artery disease (PAD) classification results for single features using multi-thresholding. The table shows area under the receiver operating characteristic curve (AUROC) values with 95% confidence intervals (CIs) across different datasets, exercise conditions, and features (photoacoustic (PA) wavelength signal, oxygen saturation (sO_2), oxygenated hemoglobin (HbO_2)).

References

[1] Isaac A Zoch et al. “Point-of-care ultrasound”. In: *Bedside Pain Management Interventions*. Springer, 2022, pp. 121–127 (cit. on p. 1).

[2] George Kossoff. “Basic physics and imaging characteristics of ultrasound”. In: *World journal of surgery* 24.2 (2000), pp. 134–142 (cit. on p. 1).

[3] Bernard Sigel. “A brief history of Doppler ultrasound in the diagnosis of peripheral vascular disease”. In: *Ultrasound in medicine & biology* 24.2 (1998), pp. 169–176 (cit. on p. 1).

[4] Nina Shevzov-Zebrun et al. “Advanced Functional Imaging: fMRI, PET, and MEG”. In: *Image-Guided Neurosurgery* (2015), pp. 63–89 (cit. on p. 1).

[5] Srirang Manohar and Daniel Razansky. “Photoacoustics: a historical review”. In: *Advances in optics and photonics* 8.4 (2016), pp. 586–617 (cit. on p. 1).

[6] Hisham Assi et al. “A review of a strategic roadmapping exercise to advance clinical translation of photoacoustic imaging: from current barriers to future adoption”. In: *Photoacoustics* 32 (2023), p. 100539 (cit. on pp. 2, 17, 53, 54, 126, 143).

[7] Soon Ho Yoon et al. “Observer variability in RECIST-based tumour burden measurements: a meta-analysis”. In: *European journal of cancer* 53 (2016), pp. 5–15 (cit. on p. 2).

[8] Zachary H Levine et al. “Uncertainties in RECIST as a measure of volume for lung nodules and liver tumors”. In: *Medical physics* 39.5 (2012), pp. 2628–2637 (cit. on p. 2).

[9] Zhenyue Chen et al. “Multimodal optoacoustic imaging: methods and contrast materials”. In: *Chemical Society Reviews* (2024) (cit. on pp. 3, 48, 49).

[10] Milou E Noltes et al. “Towards in vivo characterization of thyroid nodules suspicious for malignancy using multispectral optoacoustic tomography”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 50.9 (2023), pp. 2736–2750 (cit. on pp. 3, 53, 143, 144).

[11] Lihong V Wang and Hsin-i Wu. *Biomedical optics: principles and imaging*. John Wiley & Sons, 2007 (cit. on pp. 6–12, 18).

[12] Nicolaas Bloembergen and PS Pershan. “Light waves at the boundary of nonlinear media”. In: *Physical review* 128.2 (1962), p. 606 (cit. on p. 7).

[13] David T Delpy et al. “Pioneers in biomedical optics: special section honoring Professor Frans F. Jöbsis of Duke University”. In: *Journal of Biomedical Optics* 12.6 (2007), pp. 062101–062101 (cit. on p. 9).

[14] Frans F Jöbsis. “Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters”. In: *Science* 198.4323 (1977), pp. 1264–1267 (cit. on pp. 9, 63).

[15] Michaela Taylor-Williams et al. “Noninvasive hemoglobin sensing and imaging: optical tools for disease diagnosis”. In: *Journal of Biomedical Optics* 27.8 (2022), p. 080901 (cit. on p. 9).

[16] DH Sliney. “What is light? The visible spectrum and beyond”. In: *Eye* 30.2 (2016), pp. 222–229 (cit. on p. 9).

[17] Alexander Graham Bell. “The photophone”. In: *Science* 11 (1880), pp. 130–134 (cit. on p. 9).

[18] Guo Tang et al. “Evaluation of image reconstruction algorithms for quantitative photoacoustic tomography based on limited view data”. In: *Biomedical Spectroscopy, Microscopy, and Imaging III*. SPIE. 2024, PC130060I (cit. on p. 10).

[19] Jun Xia, Junjie Yao, and Lihong V Wang. “Photoacoustic tomography: principles and advances”. In: *Electromagnetic waves (Cambridge, Mass.)* 147 (2014), p. 1 (cit. on p. 10).

[20] Bradley E Treeby and Benjamin T Cox. “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields”. In: *Journal of biomedical optics* 15.2 (2010), pp. 021314–021314 (cit. on pp. 12, 18).

[21] Ben T Cox and Paul C Beard. “Fast calculation of pulsed photoacoustic fields in fluids using k-space methods”. In: *The Journal of the Acoustical Society of America* 117.6 (2005), pp. 3616–3627 (cit. on pp. 13, 18).

[22] Minghua Xu and Lihong V Wang. “Universal back-projection algorithm for photoacoustic computed tomography”. In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 71.1 (2005), p. 016706 (cit. on p. 13).

[23] Iben Kraglund Holfort, Fredrik Gran, and Jorgen Arendt Jensen. “Broadband minimum variance beamforming for ultrasound imaging”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 56.2 (2009), pp. 314–325 (cit. on pp. 13, 80).

[24] Giulia Matrone et al. “The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging”. In: *IEEE transactions on medical imaging* 34.4 (2014), pp. 940–949 (cit. on p. 13).

[25] Michael Jaeger et al. “Fourier reconstruction in optoacoustic imaging using truncated regularized inverse k-space interpolation”. In: *Inverse Problems* 23.6 (2007), S51 (cit. on p. 13).

[26] Andreas Hauptmann et al. “Approximate k-space models and deep learning for fast photoacoustic reconstruction”. In: *Machine Learning for Medical Image Reconstruction: First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1*. Springer. 2018, pp. 103–111 (cit. on p. 13).

[27] Minghua Xu and Lihong V Wang. “Time-domain reconstruction for thermoacoustic tomography in a spherical geometry”. In: *IEEE transactions on medical imaging* 21.7 (2002), pp. 814–822 (cit. on p. 13).

[28] Christoph Dehner et al. “A deep neural network for real-time optoacoustic image reconstruction with adjustable speed of sound”. In: *Nature Machine Intelligence* 5.10 (2023), pp. 1130–1141 (cit. on p. 13).

[29] Steven L Jacques. “Optical properties of biological tissues: a review”. In: *Physics in Medicine & Biology* 58.11 (2013), R37 (cit. on p. 14).

[30] Ben Cox et al. “Quantitative spectroscopic photoacoustic imaging: a review”. In: *Journal of biomedical optics* 17.6 (2012), pp. 061202–061202 (cit. on p. 15).

[31] Mucong Li, Yuqi Tang, and Junjie Yao. “Photoacoustic tomography of blood oxygenation: a mini review”. In: *Photoacoustics* 10 (2018), pp. 65–73 (cit. on p. 15).

[32] Roger Penrose. “On best approximate solutions of linear matrix equations”. In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 52. 1. Cambridge University Press. 1956, pp. 17–19 (cit. on p. 15).

[33] Hervé Abdi. “Singular value decomposition (SVD) and generalized singular value decomposition”. In: *Encyclopedia of measurement and statistics* 907.912 (2007), p. 44 (cit. on pp. 15, 37).

[34] SciPy Developers. *scipy.linalg.pinv — SciPy v1.11.3 Manual*. Accessed: 2025-04-02. SciPy. 2023. URL: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.linalg.pinv.html` (cit. on p. 16).

[35] Janek Gröhl et al. “SIMPA: an open-source toolkit for simulation and image processing for photonics and acoustics”. In: *Journal of biomedical optics* 27.8 (2022), pp. 083010–083010 (cit. on pp. 16, 121, 164).

[36] Niklas Joshua Holzwarth. “Spectral Unmixing in Multispectral Photoacoustic Imaging”. Thesis conducted at the German Cancer Research Center (DKFZ), supervised by Prof. Peter Bachert and Prof. Lena Maier-Hein. Bachelor’s thesis. Heidelberg, Germany: University of Heidelberg, Department of Physics and Astronomy, Oct. 2018 (cit. on pp. 16, 166).

[37] Thomas Kirchner et al. “Photoacoustic monitoring of blood oxygenation during neurosurgical interventions”. In: *Photons Plus Ultrasound: Imaging and Sensing 2019*. Vol. 10878. SPIE. 2019, pp. 14–18 (cit. on pp. 16, 166).

[38] Roman Hochuli et al. “Estimating blood oxygenation from photoacoustic images: can a simple linear spectroscopic inversion ever work?” In: *Journal of biomedical optics* 24.12 (2019), pp. 121914–121914 (cit. on p. 16).

[39] ER McFadden Jr and Harold A Lyons. “Arterial-blood gas tension in asthma”. In: *New England Journal of Medicine* 278.19 (1968), pp. 1027–1032 (cit. on p. 16).

[40] Janek Gröhl et al. “Learned spectral decoloring enables photoacoustic oximetry”. In: *Scientific reports* 11.1 (2021), p. 6565 (cit. on pp. 17, 51, 124, 138).

[41] Volker Neuschmelting et al. “Performance of a multispectral optoacoustic tomography (MSOT) system equipped with 2D vs. 3D handheld probes for potential clinical translation”. In: *Photoacoustics* 4.1 (2016), pp. 1–10 (cit. on pp. 17, 44, 86).

[42] Melanie Schellenberg. “Learning Tissue Geometries for Photoacoustic Image Analysis”. PhD thesis. 2024 (cit. on p. 18).

[43] Iván Lux. *Monte Carlo particle transport methods*. CRC press, 2018 (cit. on p. 18).

[44] Norbert N Bojarski. “The k-space formulation of the scattering problem in the time domain”. In: *The Journal of the Acoustical Society of America* 72.2 (1982), pp. 570–584 (cit. on p. 18).

[45] Bradley Treeby and Ben Cox. *k-Wave User Manual Version 1.1*. `http://www.k-wave.org/manual/k-wave_user_manual_1.1.pdf`. Department of Medical Physics and Biomedical Engineering, University College London. 2014 (cit. on p. 18).

[46] Amy M Winkler, Konstantin Maslov, and Lihong V Wang. “Noise-equivalent sensitivity of photoacoustics”. In: *Journal of biomedical optics* 18.9 (2013), pp. 097003–097003 (cit. on p. 19).

[47] Max T Rietberg et al. “Artifacts in photoacoustic imaging: Origins and mitigations”. In: *Photoacoustics* (2025), p. 100745 (cit. on p. 19).

[48] Tanja Tarvainen et al. “Bayesian image reconstruction in quantitative photoacoustic tomography”. In: *IEEE transactions on medical imaging* 32.12 (2013), pp. 2287–2298 (cit. on p. 19).

[49] Christoph Dehner et al. “Deep-learning-based electrical noise removal enables high spectral optoacoustic contrast in deep tissue”. In: *IEEE Transactions on Medical Imaging* 41.11 (2022), pp. 3182–3193 (cit. on p. 19).

[50] Jan-Hinrich Nölke et al. “Photoacoustic quantification of tissue oxygenation using conditional invertible neural networks”. In: *IEEE Transactions on Medical Imaging* (2024) (cit. on pp. 19, 51, 166).

[51] Janek Gröhl et al. “Digital twins enable full-reference quality assessment of photoacoustic image reconstructions”. In: *arXiv preprint arXiv:2505.24514* (2025) (cit. on pp. 19, 51).

[52] Enzo Ferrante and Nikos Paragios. “Slice-to-volume medical image registration: A survey”. In: *Medical image analysis* 39 (2017), pp. 101–123 (cit. on pp. 20, 48, 61, 118).

[53] Jan Modersitzki. *Numerical methods for image registration*. OUP Oxford, 2003 (cit. on pp. 20, 22).

[54] Calvin R Maurer and J Michael Fitzpatrick. “A review of medical image registration”. In: *Interactive image-guided neurosurgery* 1 (1993), pp. 17–44 (cit. on p. 20).

[55] Richard Szeliski et al. “Image alignment and stitching: A tutorial”. In: *Foundations and Trends® in Computer Graphics and Vision* 2.1 (2007), pp. 1–104 (cit. on pp. 21, 23).

[56] J Michael Fitzpatrick. “Fiducial registration error and target registration error are uncorrelated”. In: *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*. Vol. 7261. SPIE. 2009, pp. 21–32 (cit. on pp. 22, 23).

[57] J Michael Fitzpatrick. “Introduction to fiducial registration for medical applications”. In: *VISE Instructional Workshop, Summer*. 2015 (cit. on pp. 22, 23).

[58] Lang Nie et al. “Unsupervised deep image stitching: Reconstructing stitched features to images”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 6184–6197 (cit. on p. 23).

[59] Ole Vegard Solberg et al. “Freehand 3D ultrasound reconstruction algorithms—a review”. In: *Ultrasound in medicine & biology* 33.7 (2007), pp. 991–1009 (cit. on pp. 24–27, 73, 80).

[60] Mohammad Hamed Mozaffari and Won-Sook Lee. “Freehand 3-D ultrasound imaging: a systematic review”. In: *Ultrasound in medicine & biology* 43.10 (2017), pp. 2099–2124 (cit. on p. 24).

[61] Qinghua Huang and Zhaozheng Zeng. “A review on real-time 3D ultrasound imaging technology”. In: *BioMed research international* 2017.1 (2017), p. 6027029 (cit. on p. 24).

[62] Robert Rohling, Andrew Gee, and Laurence Berman. “A comparison of freehand three-dimensional ultrasound reconstruction techniques”. In: *Medical image analysis* 3.4 (1999), pp. 339–359 (cit. on pp. 24, 27).

[63] Shi Sherebrin et al. “Freehand three-dimensional ultrasound: implementation and applications”. In: *Medical Imaging 1996: Physics of Medical Imaging*. Vol. 2708. SPIE. 1996, pp. 296–303 (cit. on p. 24).

[64] Jason W Trobaugh, Darin J Trobaugh, and William D Richard. “Three-dimensional imaging with stereotactic ultrasonography”. In: *Computerized Medical Imaging and Graphics* 18.5 (1994), pp. 315–323 (cit. on p. 24).

[65] CD Barry et al. “Three-dimensional freehand ultrasound: image reconstruction and volume analysis”. In: *Ultrasound in medicine & biology* 23.8 (1997), pp. 1209–1224 (cit. on p. 26).

[66] Abigail Claus et al. “3D ultrasound-guided photoacoustic imaging to monitor the effects of suboptimal tyrosine kinase inhibitor therapy in pancreatic tumors”. In: *Frontiers in Oncology* 12 (2022), p. 915319 (cit. on pp. 27, 67).

[67] Emiel Por, Maaike van Kooten, and Vanja Sarkovic. “Nyquist–Shannon sampling theorem”. In: *Leiden University* 1.1 (2019), pp. 1–2 (cit. on pp. 28, 86).

[68] Lena Maier-Hein et al. “Metrics reloaded: recommendations for image analysis validation”. In: *Nature methods* 21.2 (2024), pp. 195–212 (cit. on pp. 29, 32–35, 150, 189).

[69] Annika Reinke et al. “Understanding metric-related pitfalls in image analysis validation”. In: *Nature methods* 21.2 (2024), pp. 182–194 (cit. on p. 29).

[70] Priya Ranganathan. "An introduction to statistics: choosing the correct statistical test". In: *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* 25.Suppl 2 (2021), S184 (cit. on p. 29).

[71] Priya Ranganathan and Nithya J Gogtay. "An introduction to statistics–data types, distributions and summarizing data". In: *Indian journal of critical care medicine: peer-reviewed, official publication of Indian Society of Critical Care Medicine* 23.Suppl 2 (2019), S169 (cit. on p. 29).

[72] Sylvia Wassertheil-Smoller and Jordan Smoller. *Biostatistics and epidemiology*. Springer, 2004 (cit. on p. 29).

[73] Priya Ranganathan and CS Pramesh. "An introduction to statistics: Understanding hypothesis testing and statistical errors". In: *Indian Journal of Critical Care Medicine: Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine* 23.Suppl 3 (2019), S230 (cit. on p. 29).

[74] Sheila F O'Brien and Qi Long Yi. "How do I interpret a confidence interval?" In: *Transfusion* 56.7 (2016), pp. 1680–1683 (cit. on p. 30).

[75] Thomas J DiCiccio and Bradley Efron. "Bootstrap confidence intervals". In: *Statistical science* 11.3 (1996), pp. 189–228 (cit. on p. 30).

[76] Philip Sedgwick. "Pearson's correlation coefficient". In: *Bmj* 345 (2012) (cit. on p. 30).

[77] Anthony J Bishara and James B Hittner. "Confidence intervals for correlations when data are not normal". In: *Behavior research methods* 49.1 (2017), pp. 294–309 (cit. on pp. 30, 31, 137).

[78] "Test for significance of Pearson's correlation coefficient". In: *International Journal of Innovative Mathematics, Statistics & Energy Policies* 6.1 (2018), pp. 11–23 (cit. on p. 31).

[79] Jan Hauke and Tomasz Kossowski. "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data". In: *Quaestiones geographicae* 30.2 (2011), pp. 87–93 (cit. on p. 31).

[80] Jerrold H Zar. "Spearman rank correlation: overview". In: *Wiley StatsRef: Statistics Reference Online* (2014) (cit. on p. 31).

[81] James A Hanley and Barbara J McNeil. "The meaning and use of the area under a receiver operating characteristic (ROC) curve." In: *Radiology* 143.1 (1982), pp. 29–36 (cit. on p. 33).

[82] John Attia. “Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests”. In: *Australian prescriber* 26.5 (2003) (cit. on p. 34).

[83] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. “Comparing images using the Hausdorff distance”. In: *IEEE Transactions on pattern analysis and machine intelligence* 15.9 (1993), pp. 850–863 (cit. on p. 35).

[84] Jiming Jiang and Thuan Nguyen. “Linear Mixed Models: Part I”. In: *Linear and Generalized Linear Mixed Models and Their Applications*. Springer, 2021, pp. 1–61 (cit. on p. 36).

[85] Hervé Abdi and Lynne J Williams. “Principal component analysis”. In: *Wiley interdisciplinary reviews: computational statistics* 2.4 (2010), pp. 433–459 (cit. on p. 36).

[86] Laurens Van der Maaten and Geoffrey Hinton. “Visualizing data using t-SNE.” In: *Journal of machine learning research* 9.11 (2008) (cit. on p. 36).

[87] Hristo Todorov, David Fournier, Susanne Gerber, et al. “Principal components analysis: theory and application to gene expression data analysis”. In: *Genomics and Computational Biology* 4.2 (2018), e100041 (cit. on p. 37).

[88] Alex Smola and SVN Vishwanathan. “Introduction to machine learning”. In: *Cambridge University, UK* 32.34 (2008), p. 2008 (cit. on p. 38).

[89] Leo Breiman. “Random forests”. In: *Machine learning* 45 (2001), pp. 5–32 (cit. on p. 38).

[90] Ahmad Taher Azar and Shereen M El-Metwally. “Decision tree classifiers for automated medical diagnosis”. In: *Neural Computing and Applications* 23 (2013), pp. 2387–2403 (cit. on p. 38).

[91] Gérard Biau and Erwan Scornet. “A random forest guided tour”. In: *Test* 25.2 (2016), pp. 197–227 (cit. on p. 38).

[92] Ian Goodfellow et al. *Deep learning*. Vol. 1. 2. MIT press Cambridge, 2016 (cit. on p. 38).

[93] Niklas Holzwarth et al. “Photoacoustic imaging for monitoring radiotherapy treatment response in head and neck tumors”. In: *Scientific Reports* 15.1 (2025), pp. 1–11 (cit. on pp. 39, 40, 55, 120–126, 130–136, 138–140, 143–145, 164, 165, 178–183, 186, 187).

[94] Athanassios Argiris et al. “Head and neck cancer”. In: *The Lancet* 371.9625 (2008), pp. 1695–1709 (cit. on p. 39).

[95] Adam Barsouk et al. “Epidemiology, risk factors, and prevention of head and neck squamous cell carcinoma”. In: *Medical Sciences* 11.2 (2023), p. 42 (cit. on p. 39).

[96] Laura QM Chow. “Head and neck cancer”. In: *New England Journal of Medicine* 382.1 (2020), pp. 60–72 (cit. on p. 39).

[97] Pierre Blanchard et al. “Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site”. In: *Radiotherapy and oncology* 100.1 (2011), pp. 33–40 (cit. on p. 39).

[98] Jordi Giralt et al. “Panitumumab plus radiotherapy versus chemoradiotherapy in patients with unresected, locally advanced squamous-cell carcinoma of the head and neck (CONCERT-2): a randomised, controlled, open-label phase 2 trial”. In: *The Lancet Oncology* 16.2 (2015), pp. 221–232 (cit. on p. 39).

[99] James PB O’Connor, Simon P Robinson, and John C Waterton. “Imaging tumour hypoxia with oxygen-enhanced MRI and BOLD MRI”. In: *The British journal of radiology* 92.1096 (2019), p. 20180642 (cit. on p. 39).

[100] Anna Bandurska-Luque et al. “FMISO-PET-based lymph node hypoxia adds to the prognostic value of tumor only hypoxia in HNSCC patients”. In: *Radiotherapy and Oncology* 130 (2019), pp. 97–103 (cit. on pp. 39, 55).

[101] Elizabeth A Eisenhauer et al. “New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)”. In: *European journal of cancer* 45.2 (2009), pp. 228–247 (cit. on p. 39).

[102] U.S. Department of Health and Human Services. *Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0*. Tech. rep. Accessed January 8, 2025. U.S. Department of Health and Human Services, Nov. 2017. URL: `https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/ctcae_v5_quick_reference_5x7.pdf` (visited on 05/16/2024) (cit. on p. 39).

[103] Sharon R Hymes, Eric A Strom, and Caroline Fife. “Radiation dermatitis: clinical presentation, pathophysiology, and treatment 2006”. In: *Journal of the American Academy of Dermatology* 54.1 (2006), pp. 28–46 (cit. on p. 40).

[104] Laurie J Rich et al. “Photoacoustic imaging as an early biomarker of radio therapeutic efficacy in head and neck cancer”. In: *Theranostics* 8.8 (2018), p. 2064 (cit. on p. 40).

[105] Giovanni Melillo. “Hypoxia and Cancer”. In: *Cancer Drug Discovery* (2014) (cit. on p. 40).

[106] Sarah E Jackson and John D Chester. "Personalised cancer medicine". In: *International journal of cancer* 137.2 (2015), pp. 262–266 (cit. on p. 40).

[107] Fred R Burlage et al. "Parotid and submandibular/sublingual salivary flow during high dose radiotherapy". In: *Radiotherapy and oncology* 61.3 (2001), pp. 271–274 (cit. on p. 40).

[108] SCH Cheng et al. "Assessment of post-radiotherapy salivary glands". In: *The British journal of radiology* 84.1001 (2011), pp. 393–402 (cit. on p. 40).

[109] Peige Song et al. "Global, regional, and national prevalence and risk factors for peripheral artery disease in 2015: an updated systematic review and analysis". In: *The Lancet Global Health* 7.8 (2019), e1020–e1030 (cit. on p. 41).

[110] Henri EJH Stoffers et al. "The prevalence of asymptomatic and unrecognized peripheral arterial occlusive disease". In: *International journal of epidemiology* 25.2 (1996), pp. 282–290 (cit. on p. 41).

[111] Heather L Gornik and Joshua A Beckman. "Peripheral arterial disease". In: *Circulation* 111.13 (2005), e169–e172 (cit. on p. 41).

[112] Anna P Träger et al. "Hybrid ultrasound and single wavelength optoacoustic imaging reveals muscle degeneration in peripheral artery disease". In: *Photoacoustics* 35 (2024), p. 100579 (cit. on pp. 41, 52, 53, 147, 158).

[113] Judith G Regensteiner et al. "Chronic changes in skeletal muscle histology and function in peripheral arterial disease." In: *Circulation* 87.2 (1993), pp. 413–421 (cit. on p. 41).

[114] Mary McGrae McDermott et al. "Leg symptoms in peripheral arterial disease: associated clinical characteristics and functional impairment". In: *Jama* 286.13 (2001), pp. 1599–1606 (cit. on pp. 41, 146).

[115] Rulon L Hardman et al. "Overview of classification systems in peripheral artery disease". In: *Seminars in interventional radiology*. Vol. 31. 04. Thieme Medical Publishers. 2014, pp. 378–388 (cit. on p. 41).

[116] Adam Behroozian and Joshua A Beckman. "Microvascular disease increases amputation in patients with peripheral artery disease". In: *Arteriosclerosis, thrombosis, and vascular biology* 40.3 (2020), pp. 534–540 (cit. on p. 41).

[117] Aaron Fenster and Donal B Downey. "3-D ultrasound imaging: A review". In: *IEEE Engineering in Medicine and Biology magazine* 15.6 (1996), pp. 41–51 (cit. on pp. 42, 45).

[118] Changyeop Lee et al. “Three-dimensional clinical handheld photoacoustic/ultrasound scanner”. In: *Photoacoustics* 18 (2020), p. 100173 (cit. on pp. 42, 45, 104, 105).

[119] Changyeop Lee, Chulhong Kim, and Byullee Park. “Review of Three-Dimensional Handheld Photoacoustic and Ultrasound Imaging Systems and Their Applications”. In: *Sensors* 23.19 (2023), p. 8149 (cit. on pp. 42, 43).

[120] Creative Commons. *Creative Commons Attribution (CC BY) license*. URL: `https://creativecommons.org/licenses/by/4.0/` (visited on 11/18/2024) (cit. on pp. 43, 62, 66, 121, 124, 130, 133, 135, 139, 140, 158, 174, 178–183, 186, 187).

[121] Laurence Mercier et al. “A review of calibration techniques for freehand 3-D ultrasound systems”. In: *Ultrasound in medicine & biology* 31.2 (2005), pp. 143–165 (cit. on pp. 43, 45).

[122] CORDIS - EU research results. *Photoacoustic/Ultrasound Mammoscopy for evaluating screening-detected abnormalities in the breast*. URL: `https://cordis.europa.eu/project/id/732411/reporting` (visited on 11/18/2024) (cit. on p. 44).

[123] Murad Omar et al. “Ultrawideband reflection-mode optoacoustic mesoscopy”. In: *Optics letters* 39.13 (2014), pp. 3911–3914 (cit. on p. 44).

[124] Sjoukje M Schoustra et al. “The Twente photoacoustic mammoscope 2: 3D vascular network visualization”. In: *Photons Plus Ultrasound: Imaging and Sensing 2019*. Vol. 10878. SPIE. 2019, pp. 63–69 (cit. on pp. 44, 105).

[125] Parker Wray et al. “Photoacoustic computed tomography of human extremities”. In: *Journal of biomedical optics* 24.2 (2019), pp. 026003–026003 (cit. on p. 44).

[126] Siyu Liu et al. “Development of a handheld volumetric photoacoustic imaging system with a central-holed 2D matrix aperture”. In: *IEEE Transactions on Biomedical Engineering* 67.9 (2020), pp. 2482–2489 (cit. on p. 44).

[127] X Luís Deán-Ben and Daniel Razansky. “Portable spherical array probe for volumetric real-time optoacoustic imaging at centimeter-scale depths”. In: *Optics express* 21.23 (2013), pp. 28062–28071 (cit. on p. 44).

[128] Seonghee Cho et al. “3D PHOVIS: 3D photoacoustic visualization studio”. In: *Photoacoustics* 18 (2020), p. 100168 (cit. on p. 44).

[129] Nikolaus Knauer, Xosé Luís Deán-Ben, and Daniel Razansky. "Spatial compounding of volumetric data enables freehand optoacoustic angiography of large-scale vascular networks". In: *IEEE Transactions on Medical Imaging* 39.4 (2019), pp. 1160–1169 (cit. on p. 45).

[130] Moein Mozaffarzadeh et al. "Motion-compensated noninvasive periodontal health monitoring using handheld and motor-based photoacoustic-ultrasound imaging systems". In: *Biomedical Optics Express* 12.3 (2021), pp. 1543–1558 (cit. on p. 45).

[131] Oliver Zettinig et al. "3D ultrasound registration-based visual servoing for neurosurgical navigation". In: *International journal of computer assisted radiology and surgery* 12 (2017), pp. 1607–1619 (cit. on p. 45).

[132] Baicheng Xing et al. "Automatic force-controlled 3D photoacoustic system for human peripheral vascular imaging". In: *Biomedical Optics Express* 14.2 (2023), pp. 987–1002 (cit. on p. 45).

[133] Thomas Kirchner et al. "Freehand photoacoustic tomography for 3D angiography using local gradient information". In: *Photons Plus Ultrasound: Imaging and Sensing 2016*. Vol. 9708. SPIE. 2016, pp. 622–632 (cit. on pp. 46, 104).

[134] Andras Lasso et al. "PLUS: open-source toolkit for ultrasound-guided intervention systems". In: *IEEE transactions on biomedical engineering* 61.10 (2014), pp. 2527–2537 (cit. on p. 46).

[135] Daohuai Jiang et al. "Hand-held free-scan 3D photoacoustic tomography with global positioning system". In: *Journal of Applied Physics* 132.7 (2022) (cit. on pp. 46, 104).

[136] Shih-Yu Sun, Matthew Gilbertson, and Brian W Anthony. "Probe localization for freehand 3D ultrasound by tracking skin features". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II 17*. Springer. 2014, pp. 365–372 (cit. on p. 46).

[137] Hedyeh Rafii-Tari, Purang Abolmaesumi, and Robert Rohling. "Panorama ultrasound for guiding epidural anesthesia: A feasibility study". In: *Information Processing in Computer-Assisted Interventions: Second International Conference, IPCAI 2011, Berlin, Germany, June 22, 2011. Proceedings 2*. Springer. 2011, pp. 179–189 (cit. on p. 46).

[138] Deeksha M Sankepalle, Brian Anthony, and Srivalleesha Mallidi. "Visual inertial odometry enabled 3D ultrasound and photoacoustic imaging". In: *Biomedical Optics Express* 14.6 (2023), pp. 2756–2772 (cit. on pp. 47, 100).

[139] Ling Dai et al. "Advancing Sensorless Freehand 3D Ultrasound Reconstruction with a Novel Coupling Pad". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, pp. 559–569 (cit. on pp. 47, 105).

[140] Niklas Holzwarth et al. "Tattoo tomography: Freehand 3D photoacoustic image reconstruction with an optical pattern". In: *International Journal of Computer Assisted Radiology and Surgery* 16 (2021), pp. 1101–1110 (cit. on pp. 47, 60–62, 64, 66, 71, 89, 93, 100, 101, 164, 174).

[141] Rüdiger Göbl et al. "Redefining ultrasound compounding: Computational sonography". In: *arXiv preprint arXiv:1811.01534* (2018) (cit. on p. 47).

[142] Jian-Feng Chen et al. "Determination of scan-plane motion using speckle decorrelation: Theoretical considerations and initial test". In: *International Journal of Imaging Systems and Technology* 8.1 (1997), pp. 38–44 (cit. on p. 47).

[143] Gunnar Farnebäck. "Two-frame motion estimation based on polynomial expansion". In: *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*. Springer. 2003, pp. 363–370 (cit. on p. 47).

[144] Wenhan Zheng et al. "Deep learning enhanced volumetric photoacoustic imaging of vasculature in human". In: *Advanced Science* 10.29 (2023), p. 2301277 (cit. on p. 47).

[145] Wolfgang Wein et al. "Three-dimensional thyroid assessment from untracked 2D ultrasound clips". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 514–523 (cit. on p. 48).

[146] Raphael Prevost et al. "3D freehand ultrasound without external tracking using deep learning". In: *Medical image analysis* 48 (2018), pp. 187–202 (cit. on pp. 48, 105).

[147] Mingyuan Luo et al. "RecON: Online learning for sensorless freehand 3D ultrasound reconstruction". In: *Medical Image Analysis* 87 (2023), p. 102810 (cit. on p. 48).

[148] Ammarah Kaderdina et al. “Using Deep Learning to Estimate Frame-to-Frame Angle Displacements in 2D Ultrasound Image Sequences of an Infant Hip”. In: *CMBES Proceedings* 45 (2023) (cit. on p. 48).

[149] Minghua Xu and Lihong V Wang. “Photoacoustic imaging in biomedicine”. In: *Review of scientific instruments* 77.4 (2006) (cit. on p. 48).

[150] Creative Commons. *Creative Commons Attribution (CC BY) license*. URL: `https://creativecommons.org/licenses/by/3.0/` (visited on 12/05/2024) (cit. on p. 49).

[151] Elena Merčep et al. “Transmission–reflection optoacoustic ultrasound (TROPUS) computed tomography of small animals”. In: *Light: Science & Applications* 8.1 (2019), p. 18 (cit. on p. 49).

[152] Yu Wang et al. “Integrated photoacoustic and fluorescence confocal microscopy”. In: *IEEE Transactions on Biomedical Engineering* 57.10 (2010), pp. 2576–2578 (cit. on p. 49).

[153] Arash Dadkhah and Shuliang Jiao. “Integrating photoacoustic microscopy, optical coherence tomography, OCT angiography, and fluorescence microscopy for multimodal imaging”. In: *Experimental Biology and Medicine* 245.4 (2020), pp. 342–347 (cit. on p. 49).

[154] Zohreh Hosseinaee, James A Tummon Simmons, and Parsin Haji Reza. “Dual-modal photoacoustic imaging and optical coherence tomography”. In: *Frontiers in Physics* 8 (2021), p. 616618 (cit. on p. 49).

[155] Marcel Gehrung et al. “Co-registration of optoacoustic tomography and magnetic resonance imaging data from murine tumour models”. In: *Photoacoustics* 18 (2020), p. 100147 (cit. on pp. 49, 119).

[156] Shuangyang Zhang et al. “MRI information-based correction and restoration of photoacoustic tomography”. In: *IEEE Transactions on Medical Imaging* 41.9 (2022), pp. 2543–2555 (cit. on p. 50).

[157] Yexing Hu et al. “Automatic image registration of optoacoustic tomography and magnetic resonance imaging based on deep learning”. In: *First Conference on Biomedical Photonics and Cross-Fusion (BPC 2022)*. Vol. 12461. SPIE. 2022, pp. 10–17 (cit. on p. 50).

[158] Raoul Kingma, Robert N Rohling, and Chris Nguan. “Registration of CT to 3D ultrasound using near-field fiducial localization: A feasibility study”. In: *Computer Aided Surgery* 16.2 (2011), pp. 54–70 (cit. on p. 50).

[159] Mattias P Heinrich et al. “MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration”. In: *Medical image analysis* 16.7 (2012), pp. 1423–1435 (cit. on p. 50).

[160] Yabo Fu et al. “Deep learning in medical image registration: a review”. In: *Physics in Medicine & Biology* 65.20 (2020), 20TR01 (cit. on p. 50).

[161] Sara Park et al. “Real-time triple-modal photoacoustic, ultrasound, and magnetic resonance fusion imaging of humans”. In: *IEEE transactions on medical imaging* 36.9 (2017), pp. 1912–1921 (cit. on pp. 50, 119, 164).

[162] Srirang Manohar et al. “Super phantoms: advanced models for testing medical imaging technologies”. In: *Communications Engineering* 3.1 (2024), p. 73 (cit. on pp. 51, 54, 119).

[163] Janek Gröhl et al. “Moving beyond simulation: data-driven quantitative photoacoustic imaging using tissue-mimicking phantoms”. In: *IEEE Transactions on Medical Imaging* 43.3 (2023), pp. 1214–1224 (cit. on p. 51).

[164] Yuqi Tang and Junjie Yao. “3D Monte Carlo simulation of light distribution in mouse brain in quantitative photoacoustic computed tomography”. In: *Quantitative imaging in medicine and surgery* 11.3 (2021), p. 1046 (cit. on p. 51).

[165] Fan Zhang et al. “Photoacoustic digital brain and deep-learning-assisted image reconstruction”. In: *Photoacoustics* 31 (2023), p. 100517 (cit. on p. 51).

[166] Ferdinand Knieling, Serene Lee, and Vasilis Ntziachristos. “A primer on current status and future opportunities of clinical optoacoustic imaging”. In: *npj Imaging* 3.1 (2025), p. 4 (cit. on pp. 51, 53).

[167] Jeongwoo Park et al. “Clinical translation of photoacoustic imaging”. In: *Nature Reviews Bioengineering* (2024), pp. 1–20 (cit. on p. 51).

[168] Amalina Binte Ebrahim Attia et al. “A review of clinical photoacoustic imaging: Current and future trends”. In: *Photoacoustics* 16 (2019), p. 100144 (cit. on p. 51).

[169] Thierry L Lefebvre et al. “The potential of photoacoustic imaging in radiation oncology”. In: *Frontiers in Oncology* 12 (2022), p. 803777 (cit. on p. 51).

[170] Li Lin and Lihong V Wang. “The emerging role of photoacoustic imaging in clinical oncology”. In: *Nature Reviews Clinical Oncology* 19.6 (2022), pp. 365–384 (cit. on p. 51).

[171] Emma Brown, Joanna Brunker, and Sarah E Bohndiek. "Photoacoustic imaging as a tool to probe the tumour microenvironment". In: *Disease models & mechanisms* 12.7 (2019), p. dmm039636 (cit. on p. 51).

[172] Christoph Becker et al. "Evaluation of cervical lymph nodes using multispectral optoacoustic tomography: a proof-of-concept study". In: *European Archives of Oto-Rhino-Laryngology* 280.10 (2023), pp. 4657–4664 (cit. on p. 51).

[173] Laurie J Rich et al. "Photoacoustic imaging as an early biomarker of radio therapeutic efficacy in head and neck cancer". In: *Theranostics* 8.8 (2018), p. 2064 (cit. on p. 52).

[174] J Vonk et al. "Multispectral optoacoustic tomography for in vivo detection of lymph node metastases in oral cancer patients using an EGFR-targeted contrast agent and intrinsic tissue contrast: a proof-of-concept study". In: *Photoacoustics* 26 (2022), p. 100362 (cit. on p. 52).

[175] Angelos Karlas et al. "Multispectral optoacoustic tomography of peripheral arterial disease based on muscle hemoglobin gradients—a pilot clinical study". In: *Annals of translational medicine* 9.1 (2021), p. 36 (cit. on p. 52).

[176] Josefine S Günther et al. "Targeting muscular hemoglobin content for classification of peripheral arterial disease by noninvasive multispectral optoacoustic tomography". In: *Cardiovascular Imaging* 16.5 (2023), pp. 719–721 (cit. on p. 52).

[177] Milenko Caranovic et al. "Derivation and validation of a non-invasive optoacoustic imaging biomarker for detection of patients with intermittent claudication". In: *Communications Medicine* 5.1 (2025), pp. 1–9 (cit. on pp. 52, 148, 159).

[178] Wonseok Choi et al. "Three-dimensional multistructural quantitative photoacoustic and US imaging of human feet in vivo". In: *Radiology* 303.2 (2022), pp. 467–473 (cit. on p. 52).

[179] NT Huynh et al. "A fast all-optical 3D photoacoustic scanner for clinical vascular imaging". In: *Nature Biomedical Engineering* (2024), pp. 1–18 (cit. on p. 52).

[180] Alexander A Oraevsky et al. "Optoacoustic imaging of oral cancer: Feasibility studies in hamster model of squamous cell carcinoma". In: *Optical Tomography and Spectroscopy of Tissue III*. Vol. 3597. SPIE. 1999, pp. 385–396 (cit. on p. 53).

[181] Jonas JM Riksen, Anton V Nikolaev, and Gijs van Soest. "Photoacoustic imaging on its way toward clinical utility: a tutorial review focusing on practical application in medicine". In: *Journal of Biomedical Optics* 28.12 (2023), pp. 121205–121205 (cit. on p. 53).

[182] Adrian P Regensburger et al. “Multispectral optoacoustic tomography enables assessment of disease activity in paediatric inflammatory bowel disease”. In: *Photoacoustics* 35 (2024), p. 100578 (cit. on p. 53).

[183] James PB O’Connor, Simon P Robinson, and John C Waterton. “Imaging tumour hypoxia with oxygen-enhanced MRI and BOLD MRI”. In: *The British journal of radiology* 92.1096 (2019), p. 20180642 (cit. on p. 55).

[184] Niklas Holzwarth et al. “Method and System for Context-Aware Photoacoustic Imaging”. European patent EP4181770B1. Granted patent. June 25, 2025. URL: https://register.epo.org/application?number=EP21745338 (cit. on pp. 56, 61, 164).

[185] Niklas Holzwarth. “3D multispectral photoacoustic tomography”. MA thesis. University of Heidelberg, 2020 (cit. on pp. 56, 60, 64, 66, 67, 70–73, 89, 93, 98, 101, 108).

[186] CDTI Horizon Europe NCP Portal: APRE. *TRL - Technology Readiness Levels*. URL: https://horizoneuropencpportal.eu/repository/71b2b25e-c6f9-4055-9c5c-1bc9f13e20d0 (visited on 12/03/2024) (cit. on p. 62).

[187] Van-Phuc Dinh et al. “Insight into the adsorption mechanisms of methylene blue and chromium (III) from aqueous solution onto pomelo fruit peel”. In: *RSC advances* 9.44 (2019), pp. 25847–25860 (cit. on p. 64).

[188] LANXESS Deutschland GmbH. *MACROLEX® Blue 3R FG - Product Data Sheet*. https://lanxess.com/. Colorant Additives Business, 5 pages. 2021 (cit. on p. 65).

[189] Stephen J Wright. *Numerical optimization*. 2006 (cit. on p. 69).

[190] Pauli Virtanen et al. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2 (cit. on p. 69).

[191] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 69).

[192] G. Bradski. “The OpenCV Library”. In: *Dr. Dobb’s Journal of Software Tools* (2000) (cit. on p. 70).

[193] Thomas S Pheiffer et al. “Model-based correction of tissue compression for tracked ultrasound in soft tissue image-guided surgery”. In: *Ultrasound in medicine & biology* 40.4 (2014), pp. 788–803 (cit. on p. 74).

[194] Salvatore Virga et al. “Use the force: deformation correction in robotic 3D ultrasound”. In: *International journal of computer assisted radiology and surgery* 13 (2018), pp. 619–627 (cit. on p. 74).

[195] Richard Beare, Bradley Lowekamp, and Ziv Yaniv. “Image segmentation, registration and characterization in R with SimpleITK”. In: *Journal of statistical software* 86 (2018), pp. 1–35 (cit. on pp. 75, 117).

[196] Fredrik Lundh. “An introduction to tkinter”. In: *URL: www. pythonware. com/library/tkinter/introduction/index. htm* (1999) (cit. on p. 78).

[197] Christoph J Bender et al. “Hardware-related biases in machine learning algorithms for photoacoustic image analysis”. In: *Photons Plus Ultrasound: Imaging and Sensing 2024*. SPIE. 2024, PC128422T (cit. on pp. 79, 158, 166).

[198] Frank R Kschischang. “The hilbert transform”. In: *University of Toronto* 83 (2006), p. 277 (cit. on p. 80).

[199] Mauro R Ruggeri and Dietmar Saupe. “Isometry-invariant Matching of Point Set Surfaces.” In: *3DOR@ Eurographics*. 2008, pp. 17–24 (cit. on p. 100).

[200] Suprosanna Shit et al. “clDice - A Novel Topology-Preserving Loss Function for Tubular Structure Segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021, pp. 16560–16569 (cit. on p. 103).

[201] Thierry L Lefebvre et al. “Performance evaluation of image co-registration methods in photoacoustic mesoscopy of the vasculature”. In: *Physics in Medicine & Biology* 69.21 (2024), p. 215007 (cit. on p. 105).

[202] Else Kröner-Fresenius-Stiftung. *Volumetrische optoakustische Bildgebung zur frühzeitigen Diagnose peripherer Gefäßerkrankungen*. Accessed: 2025-06-10. 2025. URL: `https://www.ekfs.de/wissenschaftliche-foerderung/aktuelle-foerderungen/volumetrische-optoakustische-bildgebung-zur` (cit. on pp. 106, 164).

[203] Jay B West et al. “Fiducial point placement and the accuracy of point-based, rigid body registration”. In: *Neurosurgery* 48.4 (2001), pp. 810–817 (cit. on p. 109).

[204] J Michael Fitzpatrick, Jay B West, and Calvin R Maurer. “Predicting error in rigid-body point-based registration”. In: *IEEE transactions on medical imaging* 17.5 (2002), pp. 694–702 (cit. on p. 109).

[205] James Joseph et al. “Evaluation of precision in optoacoustic tomography for preclinical imaging in living subjects”. In: *Journal of Nuclear Medicine* 58.5 (2017), pp. 807–814 (cit. on p. 110).

[206] EH Golub and C Van Loan. *Matrix Computations, Baltimore, MD: John Hopkins*. 1983 (cit. on p. 114).

[207] Kamonchanok Archawametheekul et al. “The Investigating Image Registration Accuracy and Contour Propagation for Adaptive Radiotherapy Purposes in Line with the Task Group No. 132 Recommendation”. In: *Journal of Medical Physics* 49.1 (2024), pp. 64–72 (cit. on p. 118).

[208] Athanassios Argiris et al. “Head and neck cancer”. In: *The Lancet* 371.9625 (2008), pp. 1695–1709 (cit. on p. 120).

[209] Marco Nolden et al. “The Medical Imaging Interaction Toolkit: challenges and advances: 10 years of open-source development”. In: *International journal of computer assisted radiology and surgery* 8 (2013), pp. 607–620 (cit. on pp. 121, 132).

[210] Steven L Jacques. “Optical properties of biological tissues: a review”. In: *Physics in Medicine & Biology* 58.11 (2013), R37 (cit. on p. 122).

[211] Oisín Bugter et al. “Detecting head and neck lymph node metastases with white light reflectance spectroscopy; a pilot study”. In: *Oral Oncology* 123 (2021), p. 105627 (cit. on p. 122).

[212] Kris K Dreher et al. “Unsupervised domain transfer with conditional invertible neural networks”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 770–780 (cit. on pp. 125, 126).

[213] Melanie Schellenberg et al. “Photoacoustic image synthesis with generative adversarial networks”. In: *Photoacoustics* 28 (2022), p. 100402 (cit. on pp. 125, 165).

[214] Vincent Grégoire et al. “Selection and delineation of lymph node target volumes in head and neck conformal radiotherapy. Proposal for standardizing terminology and procedure based on the surgical experience”. In: *Radiotherapy and Oncology* 56.2 (2000), pp. 135–150 (cit. on p. 131).

[215] Nicholas Schreck and Manuel Wiesenfarth. “Decomposition of Explained Variation in the Linear Mixed Model”. In: *arXiv preprint arXiv:2205.15090* (2022) (cit. on pp. 134, 135).

[216] G Gordon Steel, Trevor J McMillan, and JH2574214 Peacock. "The 5Rs of radiobiology". In: *International journal of radiation biology* 56.6 (1989), pp. 1045–1048 (cit. on pp. 134, 143).

[217] Earl F McCarthy and Donald D Van Slyke. "Diurnal variations of hemoglobin in the blood of normal men." In: (1939) (cit. on p. 137).

[218] Michael Baumann et al. "Radiation oncology in the era of precision medicine". In: *Nature Reviews Cancer* 16.4 (2016), pp. 234–249 (cit. on p. 138).

[219] Nicolas Christian et al. "Determination of tumour hypoxia with the PET tracer [18 F] EF3: improvement of the tumour-to-background ratio in a mouse tumour model". In: *European journal of nuclear medicine and molecular imaging* 34 (2007), pp. 1348–1354 (cit. on p. 138).

[220] Sebastian Zschaeck et al. "Individual patient data meta-analysis of FMISO and FAZA hypoxia PET scans from head and neck cancer patients undergoing definitive radio-chemotherapy". In: *Radiotherapy and Oncology* 149 (2020), pp. 189–196 (cit. on p. 138).

[221] Lisa M Hernández and Marcus K Taylor. "Salivary gland anatomy and physiology". In: *Salivary bioscience: foundations of interdisciplinary saliva research and applications* (2020), pp. 11–20 (cit. on p. 139).

[222] Leland McInnes, John Healy, and James Melville. "Umap: Uniform manifold approximation and projection for dimension reduction". In: *arXiv preprint arXiv: 1802.03426* (2018) (cit. on p. 142).

[223] Patricia Maria Vieten. "Deep Learning-based Semantic Segmentation of Head and Neck Photoacoustic Images". Master's thesis. Heidelberg, Germany: Department of Physics and Astronomy, University of Heidelberg, 2022 (cit. on pp. 142, 166).

[224] Patricia Vieten et al. "Deep learning-based semantic segmentation of clinically relevant tissue structures leveraging multispectral photoacoustic images". In: *Photons Plus Ultrasound: Imaging and Sensing 2022*. SPIE. 2022, PC119600P (cit. on pp. 142, 165).

[225] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pp. 203–211 (cit. on p. 142).

[226] Alexandra L Wagner et al. “Precision of handheld multispectral optoacoustic tomography for muscle imaging”. In: *Photoacoustics* 21 (2021), p. 100220 (cit. on p. 143).

[227] Wolfgang Roll et al. “Multispectral optoacoustic tomography of benign and malignant thyroid disorders: a pilot study”. In: *Journal of Nuclear Medicine* 60.10 (2019), pp. 1461–1466 (cit. on p. 144).

[228] Daniel Zips et al. “Exploratory prospective trial of hypoxia-specific PET imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer”. In: *Radiotherapy and Oncology* 105.1 (2012), pp. 21–28 (cit. on p. 144).

[229] Niklas Holzwarth et al. “Clinical tattoo tomography”. In: *Photons Plus Ultrasound: Imaging and Sensing 2023*. SPIE. 2023, PC1237930 (cit. on pp. 147, 166).

[230] Alexander Brown et al. “Detecting shortcut learning for fair medical AI using shortcut testing”. In: *Nature communications* 14.1 (2023), p. 4314 (cit. on p. 157).

[231] Marcel Knopp et al. “Shortcut learning leads to sex bias in deep learning models for photoacoustic tomography”. In: *International Journal of Computer Assisted Radiology and Surgery* (2025), pp. 1–9 (cit. on pp. 157, 158, 166).

[232] Moritz Schillinger et al. “Automatic muscle segmentation for the diagnosis of peripheral artery disease using multispectral optoacoustic tomography”. In: *Photons Plus Ultrasound: Imaging and Sensing 2025*. Vol. 13319. SPIE. 2025, pp. 304–310 (cit. on p. 159).

[233] Niklas Holzwarth et al. “Tattoo tomography: an optical pattern approach for context-aware photoacoustics”. In: *Photons Plus Ultrasound: Imaging and Sensing 2021*. Vol. 11642. SPIE. 2021, p. 1164217 (cit. on p. 164).

[234] Niklas Holzwarth et al. “Tattoo-Tomographie: Freihand-3D-Photoakustik und multimodale Bildfusion”. In: *BVM Workshop*. Springer. 2023, pp. 114–114 (cit. on p. 164).

[235] Melanie Schellenberg et al. “Semantic segmentation of multispectral photoacoustic images using deep learning”. In: *Photoacoustics* 26 (2022), p. 100341 (cit. on p. 165).

[236] Modar Kentar et al. “Spatial and temporal frequency band changes during infarct induction, infarct progression, and spreading depolarizations in the gyrencephalic brain”. In: *Frontiers in Neuroscience* 16 (2022), p. 1025967 (cit. on p. 166).

[237] Modar Kentar et al. “Mild hypothermia reduces spreading depolarizations and infarct size in a swine model”. In: *Journal of Cerebral Blood Flow & Metabolism* 43.6 (2023), pp. 999–1009 (cit. on p. 166).

[238] Edgar Santos et al. “Depth-Specific Hypoxic Responses to Spreading Depolarizations in Gyrencephalic Swine Cortex Unveiled by Photoacoustic Imaging”. In: *Translational Stroke Research* (2024), pp. 1–7 (cit. on p. 166).

[239] Sebastian Köpp. “Classification of Duchenne Muscular Dystrophy using Multispectral Photoacoustic Imaging”. Course of Studies: Computer Science (Medical Data Science). Master’s thesis. University of Applied Sciences Mannheim, Feb. 2025 (cit. on p. 166).

[240] K Thomas Robbins et al. “Neck dissection classification update: revisions proposed by the American Head and Neck Society and the American Academy of Otolaryngology–Head and Neck Surgery”. In: *Archives of otolaryngology–head & neck surgery* 128.7 (2002), pp. 751–758 (cit. on p. 180).

[241] Mahul B Amin et al. “The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging”. In: *CA: a cancer journal for clinicians* 67.2 (2017), pp. 93–99 (cit. on p. 180).

Context-aware clinical photoacoustic imaging
Ph.D. Thesis

Supervised by:	Prof. Dr. Lena Maier-Hein
Co-supervised by:	Prof. Dr. Jürgen Hesser

This work has been set using LaTeX and KOMA.

Main Font:	URW Palladio L
Math Font:	*Pazo Math*