Karlsruhe Institute of Technology

Automated Analysis Pipelines for Stem Cell-Derived Models

Zur Erlangung des akademischen Grades eines

DOKTORS DER INGENIEURWISSENSCHAFTEN (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)
angenommene
DISSERTATION

von

Luca Deininger, M.Sc.

Tag der mündlichen Prüfung:	15. Oktober 2025
Hauptreferent:	apl. Prof. Dr.-Ing. Ralf Mikut
Korreferenten:	Prof. Dr. sc. techn. ETH Jan Gerrit Korvink
	PD Dr. Dr. Daniel Hübschmann

Zusammenfassung

Stammzellbasierte Modelle bieten physiologisch relevante *in vitro*-Plattformen zur Untersuchung von Gewebeentwicklung und Krankheitsmechanismen sowohl auf Organals auch auf Organismenebene und stellen leistungsfähige Alternativen zu traditionellen humanen und tierexperimentellen Modellen dar. Sie leisten damit einen Beitrag zum Replacement im Sinne der 3R-Prinzipien für Tierversuche. Trotz ihres Potenzials erfolgt die Analyse stammzellbasierter Modelle bislang überwiegend manuell, was zu Ineffizienzen, Beobachterverzerrungen und eingeschränkter Reproduzierbarkeit führt. Zudem sind bestehende Kultivierungsprotokolle häufig wenig robust und unterliegen einer hohen Chargenvariabilität, da systematische Feedbackschleifen zur Protokolloptimierung fehlen. Diese Arbeit adressiert diese Limitationen durch die Entwicklung automatisierter Analysepipelines sowie Strategien zur Protokolloptimierung über sechs Messmodalitäten hinweg: Magnetresonanztomographie (MRT), Hellfeldmikroskopie, Epifluoreszenzmikroskopie, konfokale Live-Imaging-Mikroskopie, Einzelzell-RNA-Sequenzierung (scRNA-Seq) und Imaging Flow Cytometry. Hierzu wurden Methoden des maschinellen Lernens und Deep Learnings eingesetzt, um spezialisierte Pipelines für Segmentierung, Klassifikation und Merkmalsextraktion zu entwickeln, die eine quantitative Analyse bildgebender und einzelzellbasierter Datentypen ermöglichen.

Für stammzellbasierte Modelle auf Organebene wurde die erste automatisierte MRT-basierte Pipeline zur Segmentierung, Strukturanalyse und Qualitätskontrolle entwickelt. Für die Hellfeldmikroskopie wurde eine robuste und interpretierbare Pipeline zur Überwachung der Organoidmorphologie und -diversität implementiert, die die iterative Optimierung von Kultivierungsprotokollen unterstützt. Für Epifluoreszenzbilder wurde eine Deep-Learning-basierte Pipeline entwickelt, um Zellkerne und ventrikelartige Strukturen zu segmentieren. Darüber hinaus konnte die Machbarkeit einer *in silico* 3D-Rekonstruktion demonstriert werden. Im Bereich Einzelzellanalyse wurden Pipelines für scRNA-Seq und Imaging Flow Cytometry entwickelt, um Zellpopulationen zu quantifizieren, morphologische Merkmale zu erfassen und die Chargenvariabilität zu bestimmen. Für stammzellbasierte Modelle auf Organismusebene wurde eine neuartige Deep-Learning-Methode eingeführt, um individuelle Entwicklungstrajektorien zu verfolgen und eine frühe und späte Qualitätskontrolle zu unterstützen. Abschließend präsentiert die Arbeit allgemeine Empfehlungen zur Kultivierung, Messung, automatisierten Analyse, und Optimierung stammzellbasierter Modelle. Insgesamt zeigt sie, wie automatisierte Pipelines eine skalierbare, interpretierbare Analyse ermöglichen und die iterative Optimierung von Kultivierungs- und Messprotokollen unterstützen. Diese Beiträge schaffen die Grundlage für stabile stammzellbasierte Modelle, die eine belastbare Analyse, die Optimierung experimenteller Protokolle sowie Anwendungen in der Wissensgenerierung und Krankheitsmodellierung ermöglichen.

Abstract

Stem cell-derived models offer physiologically relevant *in vitro* platforms for studying tissue development and disease mechanisms at both the organ and organism level, serving as powerful alternatives to traditional human and animal models. In doing so, they contribute to the Replacement principle of the 3Rs (Replacement, Reduction, and Refinement) in animal research. Despite their promise, analysis of stem cell-derived models remains largely manual, leading to inefficiencies, observer bias, and limited reproducibility. In addition, cultivation protocols are often unrobust and suffer from batch-to-batch variability, as they lack systematic feedback loops for optimization. This thesis addresses these limitations by developing automated analysis pipelines and strategies for protocol optimization across six measurement modalities: magnetic resonance imaging (MRI), brightfield microscopy, epifluorescence microscopy, live-imaging-based confocal microscopy, single-cell RNA sequencing (scRNA-Seq), and imaging flow cytometry. To achieve this, machine learning and deep learning approaches were employed to build specialized pipelines for segmentation, classification, and feature extraction, enabling quantitative analysis across imaging and single-cell data types.

This thesis presents the first automated MRI-based pipeline for the segmentation, structural analysis, and quality control of organ-level stem cell-derived models. For brightfield microscopy, a robust and interpretable pipeline was implemented to monitor organoid morphology and diversity, supporting iterative protocol optimization. A deep learning-based pipeline was developed for epifluorescence microscopy images to segment nuclei and ventricle-like structures, with the feasibility of *in silico* 3D reconstruction also demonstrated. For single-cell analysis, scRNA-Seq and imaging flow cytometry pipelines were designed to quantify cell populations, assess morphological characteristics, and quantify batch variability. For organism-level stem cell-derived models, a novel deep learning-based method was introduced to monitor individual developmental trajectories and support both early- and advanced-stage quality control. Finally, the thesis presents general recommendations for the cultivation, measurement, automated analysis, and optimization of stem cell-derived models. Altogether, it demonstrates how automated pipelines enable automated, interpretable analysis and support iterative optimization of cultivation and measurement protocols. These contributions establish a foundation for stable stem cell-derived models, fostering robust analysis, protocol refinement, and applications in knowledge discovery and disease modeling.

Acknowledgements

I sincerely thank my thesis supervisor, Prof. Dr. Ralf Mikut, for his continuous support throughout my thesis, providing frequent feedback, methodological and strategic guidance, and fostering both my academic freedom and personal growth. I am also very grateful to Dr. Sabine Jung-Klawitter, my life science PI from the University of Heidelberg within the HIDSS4Health graduate school, for her invaluable feedback on biological questions. Additionally, I appreciate the insightful advice from my HIDSS4Health TAC members, Dr. Dr. Daniel Hübschmann, Dr. Daniel Schwarz, and Dr. Oya Kuseyri Hübschmann, as well as the HIDSS4Health coordinators, Dr. Ines Reinartz, Dr. Nicole Merkle, Dr. Daniel Walther, and Dr. Thomas Gerhardy, whose efforts in organizing events created a valuable network of peers facing similar challenges during their theses.

I extend my gratitude to my starting mentor, Tim, my roommate, Matthias, my imaging flow cytometry partner, Angelo, and my imaging peer group, Nils and Thomas, for their support and discussions. I also thank my AIDA seminar moderator colleague and winning AIDA table-kicker teammate, Max, as well as Lorenz, Nathalie, and Hossein, for the many campus runs, and Luca R. and Roman for their companionship during thesis writing. My gratitude extends to the entire AIDA group for the camaraderie and enriching discussions. Special thanks to Andreas Hofmann for his continuous support during my research stay abroad, and to Prof. Dr. Veit Hagenmeyer for fostering a supportive research environment. I also appreciate the IAI infrastructure team, especially Jürgen and his colleagues, for their assistance with technical matters, and the IAI secretaries, Bernadette, Manuela, Carola, and Marina, for their administrative support and positive spirit.

I am deeply grateful to Prof. Magdalena Zernicka-Goetz for welcoming me into her group at Caltech, for her valuable guidance, and for offering insight into such an inspiring research environment. I also thank Paolo for the productive collaboration, many engaging discussions, and for teaching me the world of colors.

Parts of this thesis were reformulated using ChatGPT (GPT-4o) by OpenAI. This thesis was supported by the Helmholtz Association Initiative and Networking Fund on the HAICORE@KIT partition. Moreover, I gratefully acknowledge the computational resources provided by the IAI and ML4TIME.

Finally, I wholeheartedly thank my wife, Jun, for her unwavering support during moments of frustration, my friends for standing by me in challenging times, and my family for their constant encouragement throughout this journey.

Contents

1. Introduction

> Section 1.5 about related work is partially based on: L. Deininger, P. Caldarelli, M. Zernicka-Goetz, et al. “From pixels to patterns: The AI revolution in stem cell-derived models”. Submitted to Nature Methods. 2025. [1]. The article was conceptualized and written by L.D. and P.C., with L.D. primarily responsible for the text and P.C. focusing on the figures.

1.1. Stem Cell-Derived Models

Understanding human development and disease requires physiologically relevant and scalable model systems. Stem cell-derived models are *in vitro* systems developed from pluripotent or adult stem cells that replicate key aspects of tissue development and function. They closely mimic physiological processes and provide powerful platforms for studying human and animal biology. Compared to traditional animal models like mice or zebrafish, stem cell-derived models provide a more precise representation of organ-specific processes [2–4]. Importantly, such systems contribute to the Replacement aspect of the 3Rs (Replacement, Reduction, and Refinement), which guide ethical animal research by promoting alternative methods that avoid or replace the use of animals [5].

These models serve as versatile platforms across multiple domains of biomedical research. They are widely used for knowledge discovery, enabling the study of early developmental processes and tissue morphogenesis in organs such as the brain, liver, kidney, and gut [2, 6–9]. They also support disease modeling, where patient-derived systems recapitulate disease-relevant cell types and molecular phenotypes, enabling the study of conditions such as microcephaly, Alzheimer’s, Parkinson’s disease, and various cancers [2, 10–16]. Finally, they are applied in toxicology and drug screening, where stem cell-derived models provide human-relevant platforms for compound testing, facilitating drug discovery and improving the prediction of efficacy and adverse effects [17–22].

Despite their potential, stem cell-derived models exhibit high variability across individual samples and often fail to form fully developed structures [23–27], limiting reproducibility and hindering protocol refinement. These challenges are amplified by measurement and protocol limitations, and the lack of standardized, automated tools for phenotypic analysis during or after cultivation. Automated analysis and quality assessment are therefore crucial for consistent evaluation and data-driven protocol optimization.

Stem cell-derived models vary in their complexity, ranging from systems that model individual tissues or organs to those that aim to replicate aspects of whole-organism development. This thesis focuses on one representative of each: brain organoids (Section 1.1.1) and ETiX-embryos (Section 1.1.2), respectively. While brain organoids model neural tissue organization and differentiation, ETiX-embryos simulate the integrated development of whole organisms. Together, they offer complementary insights into tissue-specific and system-level processes. Both models surpass traditional two-dimensional (2D) cultures by better capturing spatial organization, cell-cell interactions, and physiological microenvironments [28–30].

1.1.1. Brain Organoids

Brain organoids, derived from pluripotent stem cells (Figure 1.1a), mimic key features of the human brain, including its structure and cellular diversity, exemplified by the generation of major cell types such as neural progenitor cells (NPCs) and neurons [31]. These miniaturized versions of the brain are uniquely suited to studying the complexity of human brain development and associated disorders [2, 32–37]. When derived from patient-specific induced pluripotent stem cells, brain organoids enable personalized disease modeling and therapeutic discovery [38, 39].

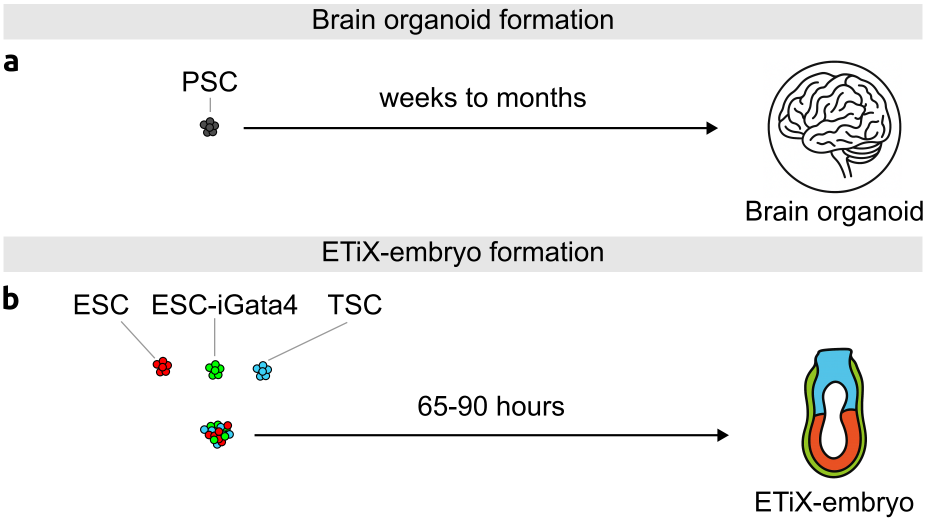


Figure 1.1: Brain organoid and ETiX-embryo formation. (a) Schematic illustration of brain organoid formation. Brain organoids are derived from pluripotent stem cells (PSCs). (b) Schematic of ETiX-embryo formation and structural organization, based on three stem cell types: embryonic stem cells (ESC), ESCs expressing inducible Gata4 (ESC-iGata4), and trophoblast stem cells (TSCs). Panel (b) is adapted from Fig. 1 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0.

Brain organoids can be categorized based on the cultivation protocol used, specifically as either *unguided* or *guided* [41]. Unguided protocols primarily rely on cell self-organization, giving rise to cerebral organoids, which are considered to model the entire brain [41, 42]. In contrast, guided protocols direct differentiation toward specific brain regions, resulting in regionalized organoids, such as forebrain organoids [41, 42]. Even though with slightly different regional specializations, forebrain and cerebral organoids exhibit significant biological similarities, both recapitulating key aspects of human brain development [43, 44].

Quantification of brain organoids is essential for effectively modeling central nervous system pathologies, particularly malformations of cortical development, which often involve changes in brain volume [45–47]. Monitoring growth and morphology during cultivation is crucial for detecting early developmental changes.

In addition, many neurological disorders are associated with disruptions in neuronal differentiation and maturation [11, 48–50]. Brain organoid substructures, such as ventricle-like regions composed of NPCs and neurons, offer valuable insights into neurodevelopment [51]. Quantification of these structures and analysis of cell type composition enable the assessment of neuronal development, with the ratio of neurons to NPCs serving as a key indicator. A higher ratio reflects more advanced developmental stages.

An illustrative application is the modeling of rare neurodevelopmental disorders, such as succinic semialdehyde dehydrogenase deficiency (SSADH). SSADH is a genetic disorder caused by mutations in the ALDH5A1 gene, leading to the accumulation of GABA and its metabolites in the brain. This causes a complex clinical presentation involving intellectual disability, ataxia, seizures, and autism spectrum traits, with no known cure [48]. Brain organoids provide a promising *in vitro* platform for modeling SSADH, enabling controlled investigation of disease mechanisms. Prior immunohistochemistry studies suggested premature neuronal differentiation might be a hallmark of the disorder SSADH [52], motivating deeper exploration of this hypothesis using quantitative automated analysis.

Brain organoid cultivation presents several challenges, in which automated quantification can play a key role in enabling iterative protocol optimization. A major limitation is the high organoid-to-organoid variability, which leads to substantial heterogeneity in cellular composition and spatial organization [53–55]. Additionally, unintended differentiation into off-target lineages further complicates reproducibility. One such undesired route is the formation of fluid-filled cavities or *cysts* [56, 57].

While many cultivation protocols focus on guiding differentiation toward specific neural lineages, they typically lack mechanisms for intrinsic quality monitoring, such as built-in criteria or feedback mechanisms for detecting suboptimal development during cultivation, e.g., detection of cystic tissue. As a result, researchers rely heavily on endpoint evaluation or manual selection, often after weeks to months of culturing. This absence of continuous, protocol-internal assessment hampers early detection of aberrant trajectories and limits opportunities for timely intervention.

1.1.2. ETiX-Embryos

ETiX-embryos are stem cell-derived embryo models engineered to mimic key developmental stages of natural embryos [26, 58]. They are generated from three stem cell types: embryonic stem cells (*E*SCs), trophoblast stem cells (*T*SCs), and ESCs transiently induced to express the transcription factor GATA4 (referred to as *iX* cells or ESC-iGata4 cells). In comparison to brain organoids, ETiX-embryos go beyond modeling individual organs by replicating early embryonic stages and the integrated processes involved in forming an entire organism (Figure 1.1b).

By the fourth day post-cell-seeding, a selection process is required to eliminate ETiX-embryos that fail to meet developmental criteria, mirroring the natural progression of embryogenesis at a comparable stage. Such selection is standard practice in organoid and stem cell-derived embryo research, where only the most promising samples are retained for further study. However, current cultivation protocols lack intrinsic quality monitoring mechanisms that could flag suboptimal development.

Automated selection is particularly necessary due to the high proportion of unsuccessfully developed ETiX-embryos. On average, only 22% of ETiX-embryos reach a well-developed state, while 78% fail to progress normally [26]. This highlights the need for improving cultivation efficiency. Moreover, the selection process remains inherently subjective, relying on individual researchers' judgment, which introduces variability across laboratories. Standardizing selection through deep-learning-based automated classification could enhance reproducibility and improve developmental success rates by ensuring a consistent and objective evaluation of ETiX-embryo viability.

1.2. Measurement Techniques

A diverse range of measurement techniques is available for characterizing stem cell-derived models, each providing unique insights into their structure, function, and developmental processes. These techniques can be broadly grouped into imaging-based methods (Sections 1.2.1-1.2.4), which are the most commonly used [59], and non-imaging techniques such as RNA sequencing (Section 1.2.5).

1.2.1. Brightfield and Fluorescence Microscopy

Brightfield and fluorescence microscopy are the most widely employed techniques for studying stem cell-derived models [59]. Brightfield microscopy uses transmitted light to illuminate transparent tissues, producing images where light-absorbing structures appear darker against a bright background. This straightforward technique is effective for assessing overall morphology and growth patterns.

Fluorescence microscopy, on the other hand, provides higher specificity by using fluorescence-labeled antibodies that bind to tissue-specific proteins, or fluorescent proteins genetically fused to target proteins, enabling detailed visualization of subcellular structures and specific features within stem cell-derived models [23, 45, 60–63]. However, as stem cell-derived systems grow larger and more opaque, full 3D imaging becomes increasingly challenging due to light scattering and limited antibody penetration. To overcome these limitations, there are two options. One solution is tissue clearing to enhance transparency, followed by light-sheet microscopy [64–67]. Yet, this method is technically demanding and relies on expensive, less-accessible equipment.

A more practical and widely accessible alternative is to use a rotary microtome to section samples into thin slices (typically $5 - 30\ \mu m$), stain them with antibodies, and image them in two dimensions. This workflow is commonly known as immunohistochemistry (IHC). To recover global tissue organization from such 2D sections, two main strategies exist. One approach involves computational 3D reconstruction and has been applied to hematoxylin-eosin-stained, paraffin-embedded serial sections [68]. Another approach uses an ultramicrotome to reconstruct large volumes from ultrathin (100 *nm*), resin-embedded sections [69, 70]. However, this approach requires specialized ultramicrotomes and labor-intensive embedding protocols. Such equipment and expertise are uncommon in most wet labs, particularly when working with fragile specimens like brain organoids. In contrast, rotary microtomes are standard in histology facilities, and IHC is well established and broadly accessible. At the start of this thesis, no related work existed about *in silico*, IHC-based, 3D reconstruction of entire brain organoids sectioned on rotary microtomes.

Among the available fluorescence microscopy systems, epifluorescence microscopes are commonly used due to their simplicity and cost-effectiveness. They illuminate the entire sample, which may result in out-of-focus fluorescence and reduced image quality. In contrast, confocal microscopes offer improved resolution through optical sectioning by using laser light and a pinhole aperture to capture fluorescence from specific focal planes, enabling 3D reconstruction capabilities [71].

Fluorescence microscopy, particularly confocal systems, is also widely used for live imaging, enabling real-time observation of dynamic cellular processes, such as cell migration, division, and differentiation in stem cell-derived models [72, 73].

1.2.2. Light-Sheet Microscopy

Light-sheet microscopy is an imaging technique that enables high-resolution, three-dimensional visualization of biological specimens. It works by illuminating the sample with a thin sheet of light perpendicular to the detection axis. This allows for fast volumetric imaging with high optical sectioning, reducing phototoxicity and photobleaching compared to techniques such as confocal microscopy [74]. Despite these advantages, the use of light-sheet microscopy remains limited due to the high cost of the required instrumentation and the complexity of sample preparation protocols, including the need for optical clearing [59].

1.2.3. Magnetic Resonance Imaging (MRI)

MRI is based on the interaction of hydrogen protons with a strong magnetic field and radiofrequency pulses, generating detailed images of tissues and structures. Although widely established in clinical and preclinical neuroscience, MRI had not been explored for imaging stem cell-derived models at the start of this thesis. Its status as the clinical gold standard for diagnosing, staging, and monitoring neurological disorders highlights its untapped potential for organoid imaging [75–77].

In contrast to many existing stem cell-derived model analysis pipelines, which are often limited to smaller structures such as intestinal organoids [78] or require destructive sample preparation [64], MRI offers the unique advantage of enabling longitudinal, non-invasive imaging of organoids over time. Compared to brightfield microscopy, the standard approach for assessing organoid size and morphology [2, 56, 79], MRI provides 3D information. T2*-weighted (T2*-w) imaging is particularly sensitive to magnetic susceptibility differences, highlighting paramagnetic substances such as neuromelanin [80], cellular debris, or calcifications. These signals can reflect biological processes such as neuronal differentiation, necrosis, or mineralization. Additionally, diffusion tensor imaging (DTI) offers functional insight into tissue architecture by capturing the directionality of water diffusion, offering insight into microstructural organization and potential fiber-like arrangements.

1.2.4. Imaging Flow Cytometry (IFC)

Imaging flow cytometry (IFC) combines traditional flow cytometry with high-resolution imaging, enabling the simultaneous analysis of both physical properties and morphological features of individual cells. Flow cytometry analyzes cell size, granularity, and fluorescence as cells pass through a laser beam, providing quantitative data for various biological applications. IFC enhances this by capturing images of cells in both brightfield and fluorescence channels, producing thousands to millions of images per experiment, allowing detailed analysis of cell morphology and function. To date, IFC has been sporadically utilized for stem cell analysis [81, 82], but has not been used to study compacted organoids or later stages of development.

1.2.5. RNA-Seq

Beyond imaging, RNA sequencing (RNA-Seq) is the most widely used method for examining stem cell-derived models [59]. RNA-Seq is a powerful technique used to analyze gene expression in biological samples, providing insights into activated and non-activated genes. RNA-seq can be divided into two main types: bulk RNA-seq and single-cell RNA-seq (scRNA-Seq). Bulk RNA-seq measures the average gene expression across a pooled population of cells, providing an overview of gene activity in the sample, while scRNA-Seq captures gene expression at the individual cell level, enabling the exploration of gene activity within distinct cell populations.

The primary output of a scRNA-Seq experiment is a matrix with dimensions $\#cells \times \#genes$, where each entry represents the transcript count for a particular gene in a given cell. For downstream analyses such as differential gene expression analysis (DGEA), data can be analyzed at the single-cell level or aggregated into pseudo-bulk profiles. In pseudo-bulk analysis, transcript counts are summed across all cells of the same group (e.g., per sample or cell type), thereby approximating bulk RNA-Seq data. This approach increases statistical robustness, reduces noise inherent to individual cells, and allows the use of established bulk RNA-Seq tools like DESeq2 [83] for DGEA.

The analysis of scRNA-Seq data involves several steps, including quality control, normalization, feature (gene) selection, data scaling, dimensionality reduction (both linear and non-linear), cell clustering, and cell annotation. Popular tools for scRNA-Seq analysis, such as Seurat [84], facilitate the semi-automated processing of data sets. After these initial steps, further downstream analyses such as differential gene expression analysis and gene set enrichment analysis provide deeper biological insights.

DGEA identifies genes with significantly different expressions between conditions, cell types, or groups, helping to understand how gene expression responds to stimuli or environmental changes. Fold change, typically reported on a log2 scale, quantifies the magnitude of expression differences (e.g., between Wildtype and Patient conditions).

Gene set enrichment analysis (GSEA) determines whether predefined gene sets, linked to biological functions, pathways, or cellular processes, show significant expression differences. GSEA often incorporates databases such as Gene Ontology (GO) [85] to interpret gene sets in the context of biological processes, enhancing the understanding of gene expression patterns and their implications for specific biological phenomena. GO is a structured framework that categorizes genes based on their associated biological processes, molecular functions, and cellular components. It is organized as a tree-like hierarchy, where broader categories branch into more specific subcategories, allowing for a systematic representation of gene functions and relationships.

1.3. Automated Image Analysis

Manual image analysis relies on human operators for inspection, segmentation, quantification, and classification, often involving hand-drawn annotations and threshold adjustments. Its accuracy and efficiency are limited by observer variability, time consumption, and subjective bias [86–88].

In contrast, automated image analysis relies on predefined algorithms, machine learning, or deep learning to extract features and make standardized, reproducible, and scalable decisions. Key tasks include image classification, which assigns labels to entire images (Section 1.3.1), image segmentation, which partitions images into meaningful regions (Section 1.3.2), image registration, which aligns images from different modalities or time points (Section 1.3.3), and feature extraction, which quantifies relevant patterns, textures, or structures (Section 1.3.4). These components work together to enable high-throughput, unbiased analysis and drive advancements in biomedical research, microscopy, and medical imaging.

1.3.1. Image Classification

Image classification is the task of assigning a label $k \in K$ to an input image $i \in I$, where I denotes the set of all images and K represents the set of predefined classes.

1.3.1.1. Classical Methods

Before the emergence of deep learning, classical image classification relied on a combination of feature extraction, discussed in Section 1.3.4, and machine learning algorithms. This approach was widely used and forms the foundation of modern deep learning methods. Support Vector Machines (SVMs) were commonly employed, leveraging high-dimensional hyperplanes to separate classes in feature-based classification tasks [89, 90]. Additionally, Random Forests and XGBoost, both decision tree-based models, improved classification performance through ensemble learning, with Random Forests reducing variance by averaging multiple trees and XGBoost enhancing accuracy through gradient boosting [91, 92].

1.3.1.2. Deep Learning

Deep learning models for image classification are dominated by convolutional neural networks (CNNs) and, more recently, vision transformers (ViTs). CNNs utilize hierarchical layers to perform convolution operations, extracting features from raw data and excelling at tasks like image classification. Notable CNN architectures include AlexNet [93], which pioneered deep network structures, MobileNet [94], designed for efficiency in mobile applications, and ResNet [95], which introduced residual connections to enable the training of very deep networks effectively. DenseNet [96] enhances gradient flow through densely connected layers, GoogLeNet [97] utilizes inception modules for multi-scale feature extraction, and ResNeXt [98] extends ResNet by incorporating grouped convolutions to improve model efficiency and performance. 3D ResNet [99], an extension of ResNet for volumetric data, is widely used in video classification and medical imaging, as it applies 3D convolutions to capture spatio-temporal features across multiple frames.

In contrast, ViTs divide images into fixed-size patches and process them as sequences of tokens through a transformer encoder, capturing global context with self-attention mechanisms [100–102]. One recent ViT, in particular, designed for video classification, is the Multiscale Vision Transformer (MViT) [102], which combines hierarchical features with ViTs. Additionally, Long Short-Term Memory (LSTM) networks, combined with a CNN as a feature extractor, are widely used in video classification to effectively model temporal dependencies [103, 104].

1.3.1.3. Evaluation Metrics

Accuracy, F1-score, and ROC AUC are widely used metrics for evaluating classification problems. Accuracy is the proportion of correct predictions over all predictions and is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{1.1}$$

Here, TP (true positives) and TN (true negatives) are cases where the model's prediction matches the actual class, while FP (false positives) and FN (false negatives) represent incorrect predictions. The terms true and false indicate whether the model's prediction is correct or not. Positive and negative refer to the presence or absence of the target feature, e.g., whether an organoid is high-quality (positive) or not (negative). While intuitive and easy to interpret, accuracy can be misleading for imbalanced datasets, where one class dominates, since it may remain high even if the model performs poorly on the minority class. The F1-score (F1, Equation 1.2) addresses this by combining precision and recall into a single harmonic mean:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \times 100 \tag{1.2}$$

Precision (Equation 1.3) measures the proportion of predicted positives that are correct, while recall (Equation 1.4) measures how well the model detects true positives:

$$precision = \frac{TP}{TP + FP} \tag{1.3}$$

$$recall = \frac{TP}{TP + FN} \tag{1.4}$$

The ROC AUC (Area Under the Receiver Operating Characteristic Curve) quantifies a model's ability to distinguish between classes across all classification thresholds. It is particularly valuable when no definitive decision threshold is specified. For an ideal classifier, Accuracy, F1-score, and ROC AUC all reach 100%.

1.3.2. Image Segmentation

Among several segmentation approaches, semantic segmentation is one of the most widely used. It assigns each image pixel (or voxel, in 3D) a label from the set $\{0, 1, \ldots, K\}$ by selecting the class with the highest predicted probability. The segmentation mask $M(p)$ is defined as:

$$M(p) = \arg\max_{k \in \{0,\ldots,K\}} P(k \mid p) \tag{1.5}$$

Here, $p \in \mathbb{Z}^n$ with $n = 2$ (i.e., $p = (x, y)$) for 2D images or $n = 3$ (i.e., $p = (x, y, z)$) for 3D images. In the special case of binary semantic segmentation ($K = 1$), the mask assigns 1 to foreground if $P(1 \mid p) > P(0 \mid p)$ and 0 to background.

1.3.2.1. Classical Methods

Classical image segmentation methods rely on mathematical and statistical principles. Among them, Otsu's method [105] is one of the most widely used global thresholding techniques. It segments an image into foreground and background by selecting an optimal threshold that minimizes intra-class variance while maximizing inter-class variance, under the assumption of a bimodal histogram. Multi-Otsu thresholding [106] extends this approach by identifying multiple threshold values to partition the image into multiple classes.

1.3.2.2. Deep Learning

Deep learning models for segmentation can be broadly categorized into CNN-based and transformer-based architectures. U-Net [107], one of the most widely used CNN-based models, features a symmetric encoder-decoder structure with skip connections. The encoder extracts hierarchical features through successive convolution and pooling operations, while the decoder restores spatial resolution using upsampling layers. Skip connections link corresponding encoder and decoder layers, integrating low-level spatial details with high-level semantic features. Numerous domain-specific adaptations of U-Net have been developed, including BeadNet [108] for bead detection, CellPose [109] for cell segmentation, and 3D U-Net [110], which extends U-Net to volumetric data.

In contrast, SegFormer represents a state-of-the-art transformer-based approach for semantic segmentation [111]. By leveraging the attention mechanisms of transformers, it captures global context more effectively than CNNs. Meanwhile, newer hybrid architectures known as foundation models, such as Segment Anything, combine CNN-based backbones with transformer-based attention mechanisms. These models offer a universal segmentation framework that generalizes across a wide range of image types and domains [112].

1.3.2.3. Evaluation Metrics

The Dice score is commonly used to quantify the performance of image segmentation methods. It is defined as twice the size of the intersection between the ground truth A and the predicted segmentation B, divided by the total number of pixels or voxels in A and B (Equation 1.6):

$$Dice\ score = \frac{2 \cdot |A \cap B|}{|A| + |B|} \tag{1.6}$$

A perfect segmentation corresponds to a Dice score of 1. In cases where both the ground truth and the prediction contain no positive class, the Dice score is mathematically undefined due to division by zero. It was therefore defined as 1.0 to reflect perfect agreement, which intuitively captures that the model correctly predicted the absence of any target structure.

1.3.3. Image Registration

Automated image registration aligns images from different modalities, time points, or perspectives by estimating geometric transformations to optimize similarity. Elastix, an open-source framework built on ITK, provides robust registration methods, including rigid, affine, and non-rigid transformations [113]. Widely used in medical imaging and microscopy, Elastix enables accurate and efficient alignment, making it a valuable tool for large-scale automated image registration.

1.3.4. Feature Extraction and Representation

Feature extraction from images plays a crucial role in quantitative analysis, with both traditional and deep learning-based approaches offering distinct advantages. Several software packages facilitate automated feature extraction from images based on a given segmentation mask. One example is PyRadiomics [114], a widely used tool for computing a broad range of features, including intensity, texture, and shape descriptors. Another commonly used library is scikit-image [115], which provides a flexible set of image processing tools and feature extraction modules that can be integrated into custom analysis pipelines.

Beyond traditional feature extraction, deep learning-based methods encode images into abstract latent spaces. For instance, DINO [116] leverages self-supervised learning to extract meaningful features from images without requiring manual annotations.

However, high-dimensional feature spaces are challenging to interpret and visualize. Principal Component Analysis (PCA) addresses this by projecting the data onto a lower-dimensional space, typically two or three dimensions, while retaining as much of the original variance as possible. This reveals feature relationships and enables the identification of meaningful clusters. Alternatively, Uniform Manifold Approximation and Projection (UMAP) offers a nonlinear embedding that preserves both local and global structure, often producing clearer separation of clusters in complex datasets [117]. UMAP has become especially popular in single-cell analyses, such as scRNA-Seq, to investigate cell populations [118].

1.4. Training and Evaluation of Deep Learning Models

Deep learning is a subset of machine learning that uses multi-layered neural networks to automatically learn hierarchical data representations. By leveraging large datasets and computational power, it extracts complex patterns without manual feature engineering, achieving state-of-the-art performance in tasks such as image classification and segmentation.

1.4.1. Training Process

The training process of deep learning models consists of four key steps. First, during the forward pass, input data is processed through the network to generate predictions. The loss function then quantifies the difference between these predictions and the ground truth. In the backward pass, gradients of the loss with respect to model parameters are computed using backpropagation. Finally, an optimizer updates the model parameters based on these gradients, the learning rate, and any regularization terms, iteratively refining the model's performance.

Training deep learning models typically requires large, well-annotated datasets. In biology, where labeled data is often scarce, transfer learning using models pretrained on large datasets, e.g., ImageNet, has become a common strategy [119]. Additionally, self-supervised learning enables models to learn useful representations from unlabeled data [116]. Both approaches are especially valuable in biomedical imaging, where annotations are costly and time-consuming.

1.4.2. Loss Functions

1.4.2.1. Image Classification

For binary image classification tasks, the binary cross-entropy (BCE) loss quantifies the discrepancy between predicted probabilities and true binary labels, guiding the optimizer to adjust model parameters for improved classification accuracy. A weighted version of BCE is defined as

$$\mathrm{L}_{\mathrm{BCE}}(y,\hat{y}) = -\frac{1}{N}\sum_{i=1}^{N}\left[w_1\, y_i\, \log(\hat{y}_i) + w_0\, (1-y_i)\, \log\bigl(1-\hat{y}_i\bigr)\right] \tag{1.7}$$

where N is the total number of samples, $y_i \in \{0,1\}$ is the ground truth label, and $\hat{y}_i$ is the predicted probability for the positive class. In the unweighted case, both classes are treated equally by setting $w_1 = w_0 = 1$. When dealing with class imbalance, the weights can be adjusted to ensure that both classes contribute proportionally to the loss. A common approach is inverse class frequency weighting, where $w_1 = \frac{N_0}{N_1}$ for the positive class and $w_0 = 1$ for the negative class, with N_1 and N_0 denoting the number of positive and negative samples, respectively.

1.4.2.2. Image Segmentation

A commonly used loss function for binary semantic segmentation combines BCE loss with Dice loss [120, 121]. It is defined as

$$\mathrm{L}(y,\hat{y}) = \alpha \cdot \mathrm{L}_{\mathrm{BCE}}(y,\hat{y}) + \beta \cdot \mathrm{L}_{\mathrm{Dice}}(y,\hat{y}) \tag{1.8}$$

where the Dice loss is given by

$$\mathrm{L}_{\mathrm{Dice}}(y,\hat{y}) = 1 - \frac{2\sum_{i=1}^{N} y_i \cdot \hat{y}}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i} \quad . \tag{1.9}$$

Here, y represents the ground truth segmentation mask and $\hat{y}$ represents the predicted segmentation mask. BCE loss provides stable pixel-wise classification but is sensitive to class imbalance, whereas Dice loss emphasizes global overlap and improves segmentation of small or underrepresented structures [120]. A 1:10 weighting ($\alpha = 1, \beta = 10$) emphasizes Dice loss to improve global overlap while preserving BCE's stability.

1.4.3. Image Augmentation

Before model training, image augmentation serves as a preprocessing step to artificially expand the size and diversity of the dataset by applying transformations such as rotation, flipping, scaling, and color jittering. These modifications enhance model robustness and reduce overfitting, particularly when training data is limited [122]. For augmentation to effectively improve model performance, it must be tailored to the specific characteristics and challenges of the given imaging task, which requires a solid understanding of the underlying problem.

To further improve generalization across different datasets, min-max normalization is commonly applied. This technique rescales each pixel intensity value $x \in X$ to a normalized range $[0, 1]$ according to:

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)}, \quad \forall x \in X \tag{1.10}$$

Here, $\min(X)$, $\max(X)$ are the minimum and maximum pixel intensity values in X, respectively.

1.4.4. Optimization

Model optimization is performed iteratively using algorithms such as Stochastic Gradient Descent [123] (SGD), which updates model parameters by computing gradients on mini-batches of data, enabling efficient training on large datasets. Training proceeds over multiple epochs, where one epoch corresponds to a complete pass through the dataset, and each epoch consists of several iterations, each based on a batch of samples. While SGD provides strong generalization properties, it can suffer from slow convergence and sensitivity to learning rate selection. To address these issues, adaptive methods such as Adam leverage adaptive learning rates and momentum to accelerate convergence [124], while AdamW improves generalization by decoupling weight decay from the gradient update, reducing the risk of overfitting [125].

To enhance generalization and mitigate overfitting, weight decay is commonly applied as a regularization technique [126, 127]. By penalizing large weights through an additional term proportional to their squared magnitude in the loss function, weight decay discourages overly complex models that fit the training data too closely, thereby improving robustness on unseen data.

1.4.5. Evaluation

Cross-validation (CV) is a technique used to assess model performance by splitting the dataset into multiple training and validation subsets. It reduces bias compared to a single train-test split. Two common types of CV are k-fold CV and leave-one-out CV.

In k-fold CV, the dataset is divided into k equally sized folds. Each fold is used once as a test set, while the remaining $k - 1$ folds are used for training (Figure 1.2). This process is repeated k times, generating split-level metrics for each fold. The results can be averaged to report overall performance, or predictions from all validation folds can be aggregated into out-of-fold predictions for the entire dataset. For deep learning, the training set within each fold is further divided into a smaller training set and a validation set, enabling model selection.

Leave-one-out CV (LOOCV) is a special case of k-fold CV where $k = N$, the number of samples. LOOCV is particularly useful for capturing sample-level variability, making it suitable for small datasets.

When constructing folds for cross-validation, stratified sampling ensures that the distribution of key features or classes in each fold reflects their proportions in the overall dataset. This approach is essential for imbalanced datasets to avoid biases for model evaluation. To properly account for confounding factors in cross-validation, data should be split in a way that prevents leakage of confounding information between training and test sets. Specifically, all samples sharing the same confounding attribute, such as originating from the same individual, batch, experimental condition, or lab, should be grouped together within a single fold. This ensures that the model is evaluated on its ability to generalize beyond those confounding factors. When only a few distinct confounding groups exist, cross-validation can be replaced or complemented by cross-testing, where the model is trained on data from one or more groups and tested on a separate, unseen group. This strategy provides an estimate of generalization performance in the presence of known sources of variability.

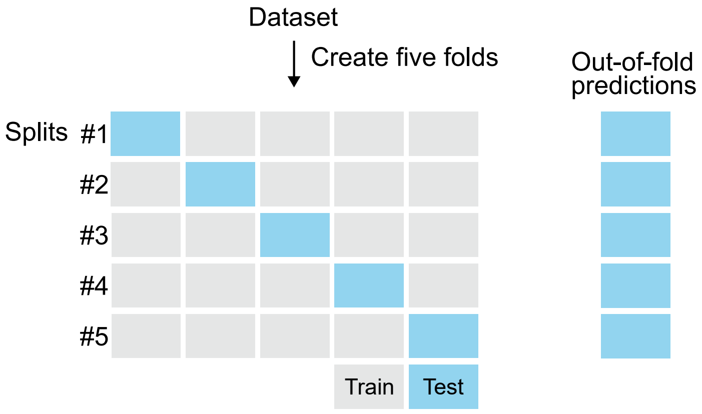


Figure 1.2: 5-fold cross-validation and out-of-fold predictions.

1.4.6. Model Explainability

Model explainability is essential for both segmentation and classification tasks, providing insights into model behavior and decision-making. In segmentation, visual inspection of results and encoder feature maps helps interpret how the model processes spatial information. For deep learning-based classification, Gradient-weighted Class Activation Mapping (Grad-CAM) [128] generates attention maps that highlight the most influential input regions, improving interpretability. In classical SVM-based classification, analyzing feature weights reveals the relative importance of individual features. Similarly, in PCA-based dimensionality reduction, examining the loading values of each feature for the principal components helps uncover feature relationships and their contributions to variance in the dataset.

1.5. Related Work

1.5.1. Manual and Semi-Automated Analysis

Image analysis of stem cell-derived models is frequently done with the open-source software Fiji [129] or CellProfiler [130], which are versatile tools supporting manual and semi-automated biological image segmentation and analysis [24, 47, 59, 131–134]. However, the growing interest in stem cell research over the past decade [135] results in an increasing amount of data and thus calls for automated analysis and quantification.

1.5.2. Automated Analysis Pipelines

1.5.2.1. Spheroids

Several analysis pipelines exist for the automated analysis of spheroids [136–144]. Spheroids are simple, typically homogeneous 3D aggregates formed by the spontaneous adhesion of one or more cell types, lacking tissue organization and differentiation potential [145]. In contrast, stem cell-derived models lead to more complex and heterogeneous structures that require more specialized analysis approaches.

1.5.2.2. Diverse Stem Cell-Derived Models

Automated analysis pipelines have been widely developed to phenotype diverse stem cell-derived organoid models, particularly using brightfield microscopy. For intestinal organoids, OrgaQuant enables quantitative assessment [146], Deep-Orga evaluates morphological features [147], and clustering methods categorize organoids phenotypically [132]. In liver and alveolar organoids, image analysis pipelines monitor growth and morphology [148], while automated frameworks extract key morphological descriptors [149]. OrganoSeg specializes in segmentation across various organoid types [150], OrgDyn captures shape dynamics of mouse mammary organoids [151], and OrganelX automates localization, segmentation, and quantification of murine liver progenitor organoids to track growth over time [152]. While these tools highlight the value of automated phenotyping, many were developed for specific imaging setups, raising concerns about their broader applicability and robustness to common variations in sample preparation and imaging conditions.

Several pipelines were also developed for fluorescence-based phenotyping. In intestinal organoids, multi-scale pipelines automate cell and substructure segmentation [78] and tracking growth dynamics [153]. Kidney organoid pipelines support segmentation [154] and differentiation status prediction [155]. Deep learning approaches enhance reproducibility by identifying morphologically similar organoids [156]. Phindr3D enables large-scale phenotypic analysis of human mammary gland acinar organoids without cell segmentation [157]. CARE automates segmentation of cell membranes and nuclei in 3D fluorescence microscopy images of human-derived cardiospheres [158].

In the context of drug screening, several organoid-focused tools have been introduced. OrganoID segments and tracks cancer organoids to extract morphology-based drug response features [159], while D-CryptO predicts both structural maturity and pharmacological response in colon organoids [160]. Cellos enables high-throughput 3D segmentation of organoids and their nuclei, supporting single-cell resolution drug response profiling in cancer organoid systems [161]. SSDquant quantifies 3D topography and viability of patient-derived glioma organoids through deep-learning-based segmentation [162].

1.5.2.3. Brain Organoids

Beyond several studies employing semi-automated analyses of gene expression data [163–165] and a single approach applied to light-sheet imaging [64], several image-based pipelines have been developed for analyzing brain organoids using brightfield and fluorescence microscopy.

For brightfield imaging, MOrgAna is an automated analysis pipeline that combines multilayer perceptron models and logistic regression to perform segmentation and morphology-based quantification of brain organoids in brightfield microscopy images [166]. It features a graphical user interface to enable broad accessibility, allowing users without expert knowledge to perform analyses efficiently. Another study also designed a segmentation pipeline for brightfield images [167]. Additionally, clustering has been applied to characterize and differentiate the morphological development of brain organoids at different stages using segmented brightfield images [168].

In the fluorescence domain, an early semi-automated pipeline applies thresholding and traditional image-processing steps to count cells and detect ventricle-like regions, though it does not specify whether epifluorescence or confocal imaging was used [169]. More recently, a deep learning-based tool was introduced to perform classification, segmentation, and object detection on both brightfield and fluorescence images of brain organoids, enabling morphological feature extraction to support developmental analysis [170].

1.5.2.4. Embryo Models

Several automated analysis pipelines have been developed for embryo models, though the majority focus on non-stem cell-derived models. These include AI-driven approaches for *in vitro* fertilization success prediction, phenotypic defect detection and staging in zebrafish, structural quantification in frog and mouse embryos, and high-throughput screening applications [171–176].

For stem cell-derived embryo models, deep learning has been applied to automate classification and analysis. One study developed a deep learning-based classifier that categorizes human blastoids into five quality groups using brightfield images for high-throughput evaluation [177]. However, at the start of this thesis, no automated analysis pipelines had been developed for embryo models of later developmental stages, such as ETiX-embryos. Nonetheless, advancements in automated analysis for embryo models and other stem cell-derived systems underscore the potential of deep learning for automating classification and quantitative analysis of complex, organism-level models like ETiX-embryos.

1.6. Open Questions

There are several open questions in the domain of automated analysis pipelines for stem cell-derived models that need to be addressed:

1. While several organoid analysis pipelines have been developed for brightfield microscopy, magnetic resonance imaging (MRI) remained unexplored for organoid analysis, including brain organoids, at the start of this thesis. Given its non-invasive nature and ability to capture internal structures in 3D without the need for labeling, MRI offers unique advantages. How can an automated MRI-based analysis pipeline be designed to extract brain organoid structure and assess organoid quality? Furthermore, what unique features does an MRI-based analysis pipeline provide compared to automated organoid analysis using brightfield microscopy?

2. A wide range of organoid analysis pipelines for brightfield microscopy images exist, yet their robustness to imaging artifacts, such as light reflections, shadows, and color variations from the culture medium, varies. This leads to inaccurate monitoring of growth and morphology, limits the transferability of methods to other laboratories, and restricts insights into developmental dynamics. How can a segmentation and analysis pipeline be implemented to ensure robustness and explainability for brain organoid size monitoring, assessment of organoid diversity, and iterative protocol optimization?

3. Compared to brightfield microscopy, epifluorescence microscopy provides deeper structural insights into brain organoids while being less complex than light-sheet imaging. How can an automated analysis pipeline be developed to quantify nuclei and ventricle-like structures from epifluorescence microscopy images? Additionally, is *in silico* volumetric reconstruction from serial epifluorescence images a feasible approach to capture global organoid architecture, and what challenges must be addressed?

4. Compared to organoid-level analyses, cell-level analysis provides additional insights into the presence and characteristics of distinct cell populations. How can automated

analysis pipelines for single-cell RNA sequencing and imaging flow cytometry be leveraged for brain organoid analysis and quality control, and what unique advantages do these approaches offer in comparison to organoid-level analyses?

5. While previous analysis pipelines have focused on phenotyping and quality control in organ-level stem cell-derived models, similar challenges arise in systems of whole-organism development. How can these pipeline elements be adapted to support automated quality monitoring and iterative protocol optimization in more complex, organism-level stem cell-based systems?

6. Experimental protocols for stem cell-derived models have so far lacked systematic feedback from automated analysis pipelines. What feedback from automated analysis pipelines can be used to enhance the cultivation and measurement of these systems? And what conclusions and recommendations can be derived for advancing their automated analysis?

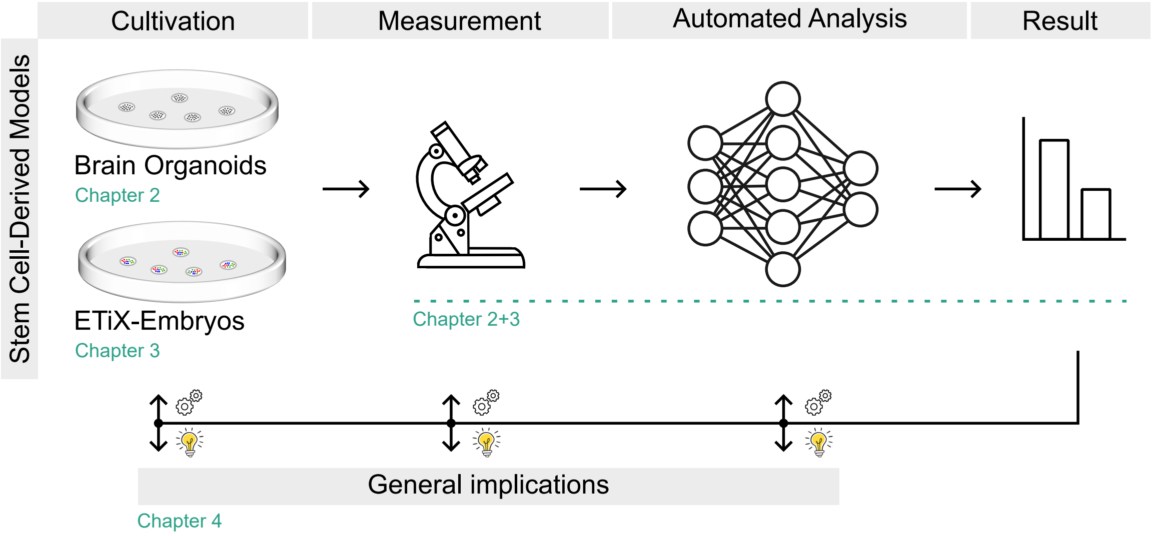


Figure 1.3: Thesis outline.

1.7. Objectives and Thesis Outline

Based on the previously outlined open questions, the primary objectives of this thesis are:

1. Develop an automated analysis pipeline for brain organoids using MRI data to monitor organoid morphology and quality, while comparing its performance with brightfield imaging for the same tasks.

2. Establish a robust and explainable automated analysis pipeline for brain organoids using brightfield microscopy data to monitor organoid size and morphological

characteristics while demonstrating how automated methods can facilitate iterative protocol optimization.

3. Design an automated analysis pipeline for brain organoids using epifluorescence microscopy images to quantify nuclei and ventricle-like structures. Additionally, investigate and evaluate the feasibility of registration-based volumetric reconstruction from serial epifluorescence immunohistochemistry sections of brain organoids.
4. Explore the potential of automated scRNA-Seq data analysis and imaging flow cytometry for brain organoids, comparing their capabilities to image-based modalities and evaluating their roles in protocol optimization.
5. Implement an automated analysis pipeline incorporating a novel deep learning approach for quality monitoring, variability assessment, and cultivation improvement of organism-level stem cell-derived models at both early and advanced developmental stages based on live-imaging confocal microscopy.
6. Generate insights into the broader implications of automated quantification, measurement, and protocol optimization for stem cell-derived models, with an emphasis on recommendations for robust, comprehensive, and targeted analyses.

Figure 1.3 provides an overview of how the contributions of each chapter integrate into the thesis. Chapter 2 introduces analysis pipelines developed for brain organoids, which represent organ-level stem cell models, across multiple data modalities. Building on these foundations, Chapter 3 focuses on adapting analysis pipelines for ETiX-embryos, which represent organism-level stem cell models. Finally, Chapter 4 broadens the perspective by delineating general implications for the cultivation, measurement, automated analysis, and optimization of experimental protocols of stem cell-derived models.

2. Analysis Pipelines for Brain Organoids

As part of this work, automated analysis pipelines for brain organoids were developed, investigated, and evaluated across several data modalities, addressing automated quantification of organoid size, growth, quality, diversity, cellular substructures, and cell populations. These modalities include magnetic resonance imaging (MRI), as detailed in Section 2.1, brightfield microscopy, discussed in Section 2.2, and epifluorescence microscopy, covered in Section 2.3. Additionally, cell-level analysis was explored in Section 2.4, which includes single-cell RNA-Seq, described in Section 2.4.1, and imaging flow cytometry, outlined in Section 2.4.2.

Throughout this study, all sections focus on cerebral organoids, except for Section 2.2, which examines forebrain organoids — a closely related type of brain organoid with many shared characteristics, as explained in Section 1.1.1.

2.1. Magnetic Resonance Imaging (MRI)

> This section is based on work presented in L. Deininger, S. Jung-Klawitter, R. Mikut, et al. “An AI-based segmentation and analysis pipeline for high-field MR monitoring of cerebral organoids”. In: *Scientific Reports*, Vol. 13, No. 1 (Dec. 2023). ISSN: 2045-2322. DOI: `10.1038/s41598-023-48343-7`. [178]. S.J-K. conducted organoid cultivation, while D.S. handled the magnetic resonance imaging. L.D. was responsible for data processing, data analysis, and the development of the automated analysis pipeline.

While automated analysis pipelines for cerebral organoids have been established for brightfield and fluorescence microscopy, magnetic resonance imaging (MRI) remained an untapped modality at the start of this thesis. Leveraging its ability to capture internal 3D structures non-invasively and without labeling, this section explores how MRI can be used for automated analysis and what distinct advantages it offers over brightfield-based methods.

In this work, the first automated pipeline for analyzing cerebral organoids in MRI is introduced, using a neural network-based approach to extract organoid volume and structural features. Specifically, three crucial tasks for cerebral organoid monitoring and quality assessment are addressed: (i) organoid segmentation (Section 2.1.4.1), (ii) global cysticity classification (Section 2.1.4.2), and (iii) local cyst segmentation (Section 2.1.4.3). Finally, its performance is compared to those of brightfield-based analysis pipelines (Section 2.1.4.5).

2.1.1. Sample Preparation and MRI

Sample preparation involved transferring the organoids into medium-containing Eppendorf tubes and positioning them in a custom holder, enabling the simultaneous imaging of three organoids (Figure 2.1a-c). This setup optimized the use of the MRI scanner’s imaging volume, ensured adequate spacing to prevent signal interference or artifacts, and reduced total imaging time by allowing parallel acquisition. The imaging medium (organoid differentiation medium) was chosen to maintain organoid viability and minimize metabolic stress during MRI. In addition, the medium was selected for its low magnetic susceptibility and minimal background signal, thereby reducing artifacts and ensuring optimal image contrast.

MRI was performed using a high-field 9.4 Tesla horizontal bore small animal scanner (BioSpec 94/20 USR, Bruker BioSpin GmbH, Ettlingen, Germany). Compared to conventional clinical MRI systems (1–3 Tesla) or preclinical scanners with field strengths up to 7 Tesla, this high-field setup offers significantly improved spatial resolution and signal-to-noise ratio [179, 180]. Two MRI sequences were recorded: a 3D T2*-w sequence with a resolution of $387 \times 100 \times 75$ px and an isotropic voxel size of 80 μm, and a 2D Diffusion Tensor Imaging (DTI)-spin echo sequence with a resolution of $309 \times 100 \times 12$ px and 100 μm in-plane resolution. Further details regarding MRI parameters can be found in [178].

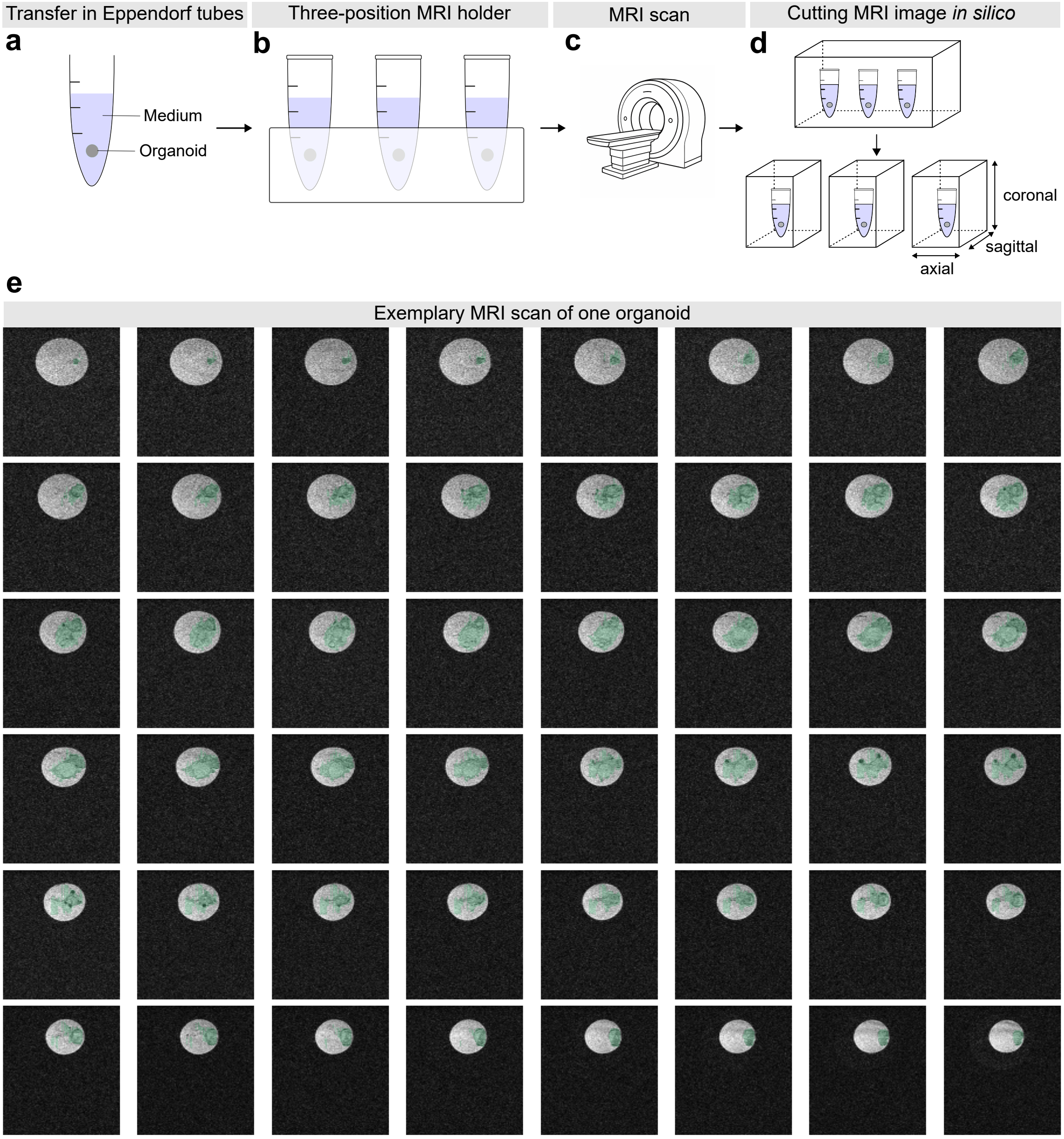


Figure 2.1: Experimental setup, data preparation, and data. (a) Transfer of organoids into medium-containing Eppendorf tubes. (b-c) Placement of three Eppendorf tubes in a holder for MRI. (d) After MRI, the MRI images were cut *in silico* to derive one image per organoid (e) 2D coronal planes of the MRI image (T2*-w sequence) of one organoid. Ground truth organoid segmentation as a green transparent overlay showing organoid localization. The first 16 and last 11 planes were omitted for space reasons. Derivative of Fig. S6 by Deininger et al. [178] licensed under CC BY 4.0; added details and added (e).

Nine control organoids, which did not undergo MRI, served as handling controls to assess any potential effects of the imaging process itself. Ensuring that MRI does not adversely affect organoid metabolism or viability is crucial, as any such influence could distort the biological interpretation of the results. Blood gas analysis of the culture medium before and after imaging confirmed that the MRI procedure had no specific negative impact on organoid health (Figure A.7). The handling controls thereby validated that the imaging setup and scanning conditions preserved physiological stability during MRI.

2.1.2. Dataset

Two MRI datasets were generated for this work, designated as Dataset A ($Data_A$) and Dataset B ($Data_B$). The datasets were acquired in separate experimental runs but followed the same protocols. Each dataset consists of scans of nine wildtype cerebral organoids taken at various time points over 64 days, resulting in a total of 45 samples (Figure 2.2). Unless specified otherwise, figures and text refer to results from $Data_A$. The results from $Data_B$ are included to validate the methodology and to demonstrate the performance of the organoid and local cyst segmentation models trained on $Data_A$ and tested on $Data_B$, and vice versa.

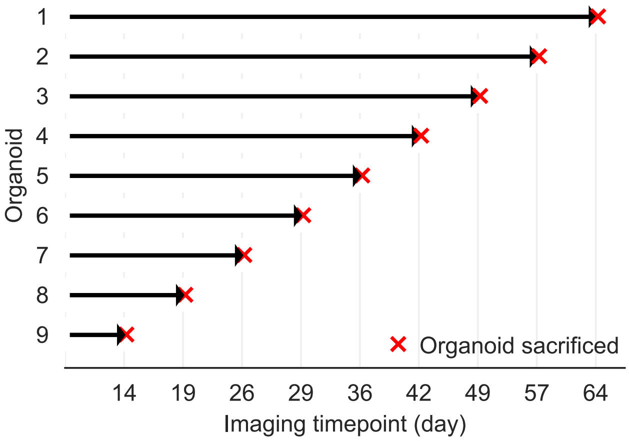


Figure 2.2: Data acquisition. Nine organoids were imaged at multiple time points over 64 days. After each imaging session, one organoid was sacrificed for additional experiments. This sequential sacrifice resulted in a total of 45 MRI samples, with organoid 1 imaged 9 times, organoid 2 imaged 8 times, and so on, down to organoid 9, which was imaged only once.

Due to the chosen imaging setup, three organoids appeared within a single MRI image. As a first preprocessing step, the image was separated *in silico* into three equal sections to isolate each organoid for downstream analysis (Figure 2.1d,e). Each T2*-w organoid image measured 129 × 100 × 75 px (axial × sagittal × coronal), and each DTI image measured 103 × 100 × 12 px (axial × sagittal × coronal).

Annotations were generated to support the development and evaluation of the image analysis methods presented in this work. One expert annotated both MRI datasets, including organoid and local cyst segmentations based on the T2*-w sequence, and global cysticity classification labels. All MRI images and corresponding annotations are publicly available on Zenodo (`https://zenodo.org/record/7805426`). The code used to reproduce the results is accessible on GitHub (`https://github.com/deiluca/cerebral_organoid_quant_mri`).

2.1.3. Methods

2.1.3.1. Organoid Segmentation

To assess organoid volume and morphology, images from the T2*-w sequence were utilized. Binary semantic segmentation was employed to classify each voxel as either organoid (foreground) or non-organoid (background, Equation 1.5). To enhance model generalization across different datasets, min-max normalization was applied to each 3D scan (Equation 1.10). Three methods were evaluated for MRI organoid segmentation: Multi-Otsu thresholding, 2D U-Net, and 3D U-Net.

Multi-Otsu thresholding was applied on each MRI image rather than per coronal plane, due to uncertainty about the presence of the organoid in every plane, as the organoid occupies only small portions of the images (Figure 2.1e). Using the scikit-image Python package [115], MRI images were segmented into three classes: MRI background, Eppendorf tube including medium, and organoid. To enable evaluation and comparison with binary segmentation, the organoid class prediction was used. The MRI background showed low image intensities, the Eppendorf tube high image intensities, and the organoid intermediate image intensities; therefore, Multi-Otsu's threshold assigned the intermediate intensity range to the organoid.

For the 2D U-Net [107], data augmentation included random rotations by 0°, 90°, 180°, or 270°, applied with a probability of 0.75. Training and evaluation were performed on 2D slices extracted along the coronal axis of each MRI image (Figure 2.1d), using the implementation from [181]. The model was trained for 200 epochs with the Adam optimizer (learning rate = 1×10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 1×10^{-3}), using a batch size of 1.

The 3D U-Net [110] was trained on 3D volumes using the implementation from [182], also with the Adam optimizer (learning rate = 1×10^{-3}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 1×10^{-7}), over 2,000 iterations and a batch size of 1.

2.1.3.2. Global Cysticity Classification

Global cysticity classification aims at determining the overall organoid cysticity: cystic (low-quality) or non-cystic (high-quality). To provide a reference ground truth based on the T2*-w sequence, an organoid was categorized as cystic if at least one cystic structure was detected within the organoid. Otherwise, it was categorized as non-cystic.

For automatic classification in MRI, the metric *Compactness* was constructed, which serves as an environment-based estimator of organoid cysticity (Equation 2.1). It is based on the idea that cysts are filled with a fluid similar to the medium. Therefore, the more similar the organoid intensities (Equation 2.5) are to the medium intensities (Equation 2.6), the more cystic the organoid is (Figure 2.3a).

$$\text{Compactness} := \left|\mu(\mathbf{v}_{\text{org}}) - \mu(\mathbf{v}_{\text{medium}})\right| \tag{2.1}$$

$$\mu(X) = \frac{1}{|X|}\sum_{x\in X} x \tag{2.2}$$

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{otherwise} \end{cases} \tag{2.3}$$

$$A \setminus B = \{x \in A \mid x \notin B\} \tag{2.4}$$

$$\mathbf{v}_{\text{org}} = \{\text{intensities of organoid voxels}\} \tag{2.5}$$

$$\mathbf{v}_{\text{medium}} = \{\text{intensities of medium voxels}\} \setminus \mathbf{v}_{\text{org}} \tag{2.6}$$

While $\mathbf{v}_{\text{org}}$ was derived from the ground truth organoid segmentations, $\mathbf{v}_{\text{medium}}$ was determined by applying Otsu's threshold [105] in 2D across all coronal planes containing organoids, excluding organoid regions based on the ground truth segmentations (Figure 2.3b). The first and last organoid-containing coronal planes were discarded to filter artifacts caused by noisy medium intensities.

To compare *Compactness* with a deep learning model, a 3D ResNet18 [183] was used. The model was trained with stochastic gradient descent (learning rate = 1×10^{-3}) for 150 epochs with a batch size of 1. A binary cross-entropy loss function, weighted by inverse class frequencies, was employed to ensure an unbiased training process.

To investigate not only the organoid quality but also the tissue characteristics of cystic and non-cystic organoids, diffusion tensor imaging (DTI) was employed. DTI enables the extraction of quantitative parameter maps—such as the mean diffusivity (Trace), fractional anisotropy (FA), and the first, second, and third eigenvalues of the diffusion tensor—which provide insight into the microstructural organization of the tissue. These parameters were calculated for each organoid, and a two-sided t-test was used to assess statistically significant differences in average diffusion values between cystic and non-cystic groups. To account for multiple comparisons, p-values were adjusted using the Holm-Šídák method.

2.1.3.3. Local Cyst Segmentation

Local cyst segmentation aims at localizing cystic regions within brain organoids. For this task, the T2*-w sequence was used, and cysts were manually annotated. Due to the limited spatial resolution, small cysts lie near the detection limit of the MRI and are difficult to distinguish from noise or partial volume effects, making reliable annotation challenging even for human experts and prone to label noise. To ensure the quality and consistency of the segmentation ground truth, organoids with fewer than 1,000 cyst voxels ($0.51, \text{mm}^3$) were excluded from the dataset, resulting in a final set of 34 samples. This exclusion is

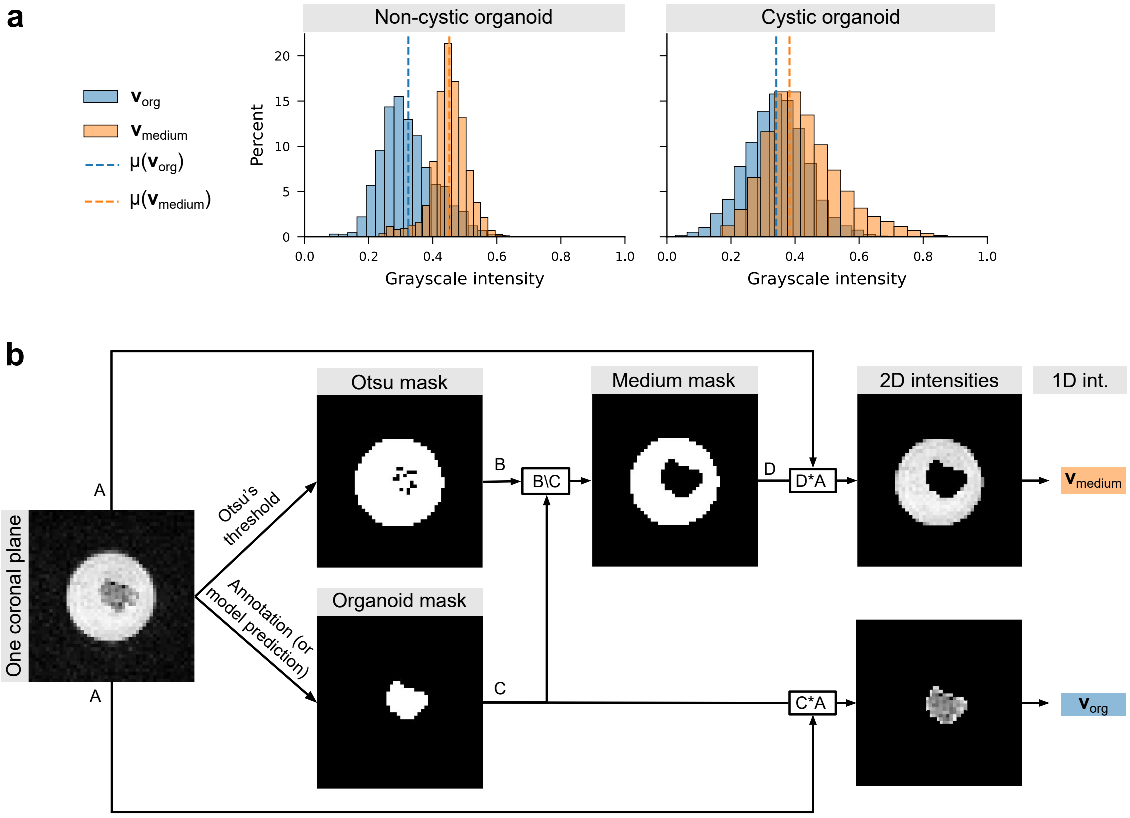


Figure 2.3: Concept of global cysticity classification in MRI. (a) Histograms depicting the grayscale intensity distributions of voxels corresponding to organoid tissue ($\mathbf{v}_{\mathrm{org}}$) and surrounding medium ($\mathbf{v}_{\mathrm{medium}}$) in both non-cystic and cystic organoids. (b) Otsu's mask, organoid location, and medium mask are binary masks. The white pixels of the medium mask belong to the medium. This example is based on Organoid 1 (day 14), coronal plane 60. To determine the medium intensities for one organoid, this procedure is applied to all organoid-containing coronal planes from the 3D image. For better visibility in this figure, the coronal plane shown was cut to the Eppendorf tube boundaries. Derivative of Fig. S7 by Deininger et al. [178] licensed under CC BY 4.0, with added panel (a) and modified layout in (b).

unproblematic from a quality assessment perspective, as organoids with only small or ambiguous cysts typically still exhibit good overall structural integrity and would not be classified as low-quality based on their cysticity alone.

For segmentation, a 3D U-Net [110] was trained and evaluated as for organoid segmentation but with 5,000 training iterations. Local cyst segmentation was performed for MRI images only, because the annotation of cysts in brightfield images was not feasible due to the limited spatial visibility of cysts in the 2D images (Figure A.5).

2.1.3.4. Model Evaluation

For the evaluation of organoid segmentation and local cyst segmentation, the Dice score was used (Equation 1.6). To obtain an unbiased estimate of model performance, organoid-wise leave-one-out cross-validation (LOOCV) was employed. In each of the nine LOOCV splits, all images from a single organoid were reserved for model testing, while the remaining images from all other organoids were used for model training and validation. Within each LOOCV split, the training set was further divided into 80% training and 20% validation data, assigned randomly at the image level rather than organoid-wise to ensure a more stable number of images in both subsets. For the assessment of global cysticity classification, the area under the Receiver Operating Characteristic curve (ROC AUC) was used to evaluate multiple thresholds. Additionally, organoid-wise LOOCV was applied to assess the performance of the 3D ResNet18 model for global cysticity classification.

2.1.4. Results

2.1.4.1. Organoid Segmentation

Organoid segmentation is essential to automatically extract features like organoid volume or structure. Neither thresholding nor using a 2D U-Net produced satisfactory results for MRI-based organoid segmentation (Figure 2.4). The 3D U-Net achieved superior performance with a Dice score of 0.92 ± 0.06 (mean ± SD, Figures 2.4, 2.5).

Although the 3D U-Net performed highly accurately overall, challenging samples were analyzed to identify its weaknesses. The model performed worst on Organoid 3 on day 36, with a Dice score of 0.59. In this case, the disruption of one or more cystic structures led to a reduced overall volume (Figure A.3) and fragmented the organoid into multiple pieces (Figure 2.5c). These fragments adhered to the Eppendorf tube wall, causing parts of the organoid's border to blur with the MRI background. As a unique biological outlier in the dataset, this sample posed a learning challenge for the model. In contrast, the analysis of other samples demonstrated reliable detection (Figure 2.5d-e).

For application to larger-scale experiments, it is essential that both model training and, especially, inference times remain within a practical range. The 3D U-Net completes training for MRI organoid segmentation in under an hour, with inference times averaging approximately two seconds per sample (Table A.1).

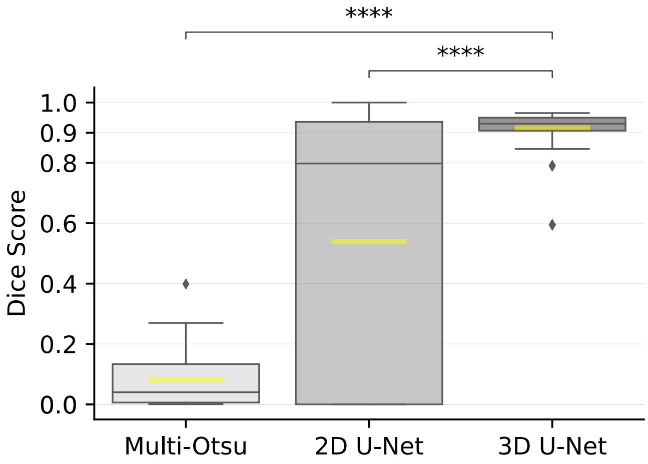


Figure 2.4: Organoid segmentation performance of Multi-Otsu's threshold, 2D U-Net, and 3D U-Net. The boxplots show the Dice scores on the LOOCV test sets: 45 values for Multi-Otsu thresholding and the 3D U-Net (one per sample), and 3,375 values for the 2D U-Net (# samples × # coronal planes). Each box represents the interquartile range (IQR), with the median shown as a black line and the mean indicated by a yellow line. Whiskers extend to the minimum and maximum values within 1.5 times the IQR. ****$p<0.0001$, two-sided Welch's t-test.

2.1.4.2. Global Cysticity Classification

Cyst formation is an undesired process during cerebral organoid cultivation (Section 1.1.1). Thus, accurately determining organoid cysticity can serve as a quality control tool. Qualitatively, non-cystic and cystic organoids show different morphologies in MRI and brightfield microscopy (Figure A.5). While segmentation performance was evaluated using the Dice score, the cysticity classification was assessed using the ROC AUC. For global cysticity classification, the environment-based metric *Compactness* achieved an ROC AUC of 0.98 (Table 2.1, Figure 2.6), thus highlighting its role as a reliable quality control tool. Separating cystic and non-cystic samples using either a 3D ResNet or using their mean intensities resulted in lower ROC AUCs (Table 2.1). Different diffusion tensor imaging (DTI) maps can be used to distinguish cystic and non-cystic organoids using the organoid mean intensity, but at a lower ROC AUC compared to using *Compactness* on the T2*-w sequence (Table 2.1).

Compared to the T2*-w sequence, DTI enables the investigation of tissue properties such as diffusion characteristics. DTI showed a significantly higher average diffusion in cystic organoids compared to non-cystic organoids (Figure 2.7a). As shown in Figure 2.7b-c, regions with increased diffusion correspond to cystic areas. This difference most likely reflects the higher fluid content of cystic organoids, as fluid-filled structures allow for less restricted water diffusion.

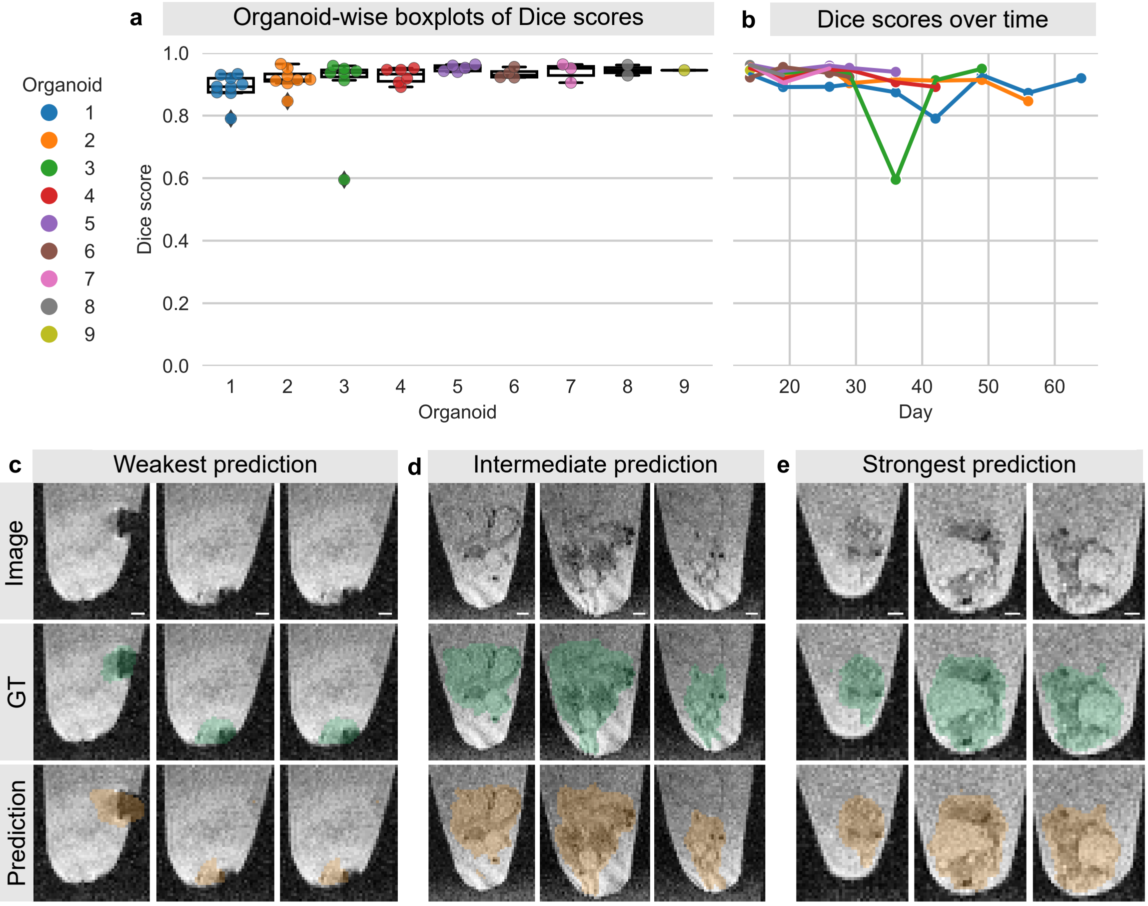


Figure 2.5: 3D organoid segmentation in MRI. (a-b), Model performance. (c-e), Selected sagittal planes. (c), Organoid 3 (day 36): Dice score of 0.59. (d), Organoid 2 (day 42): Dice score of 0.91. (e), Organoid 5 (day 26): Dice score of 0.95. Image: original image, GT: Image with ground truth organoid location (green), Prediction: image with predicted organoid location (orange). Selected sagittal planes (left to right): (c), 50, 47, 44; (d), 58, 50, 40; (e), 52, 40, 34. For better visibility of the organoids and to investigate the detailed differences between ground truth and model prediction, the images were cut to the organoid location. Scale bar: 400 μm. Derivative of Fig. 1 in Deininger et al. [178], licensed under CC BY 4.0, with adapted panels (a,b) and modified layout.

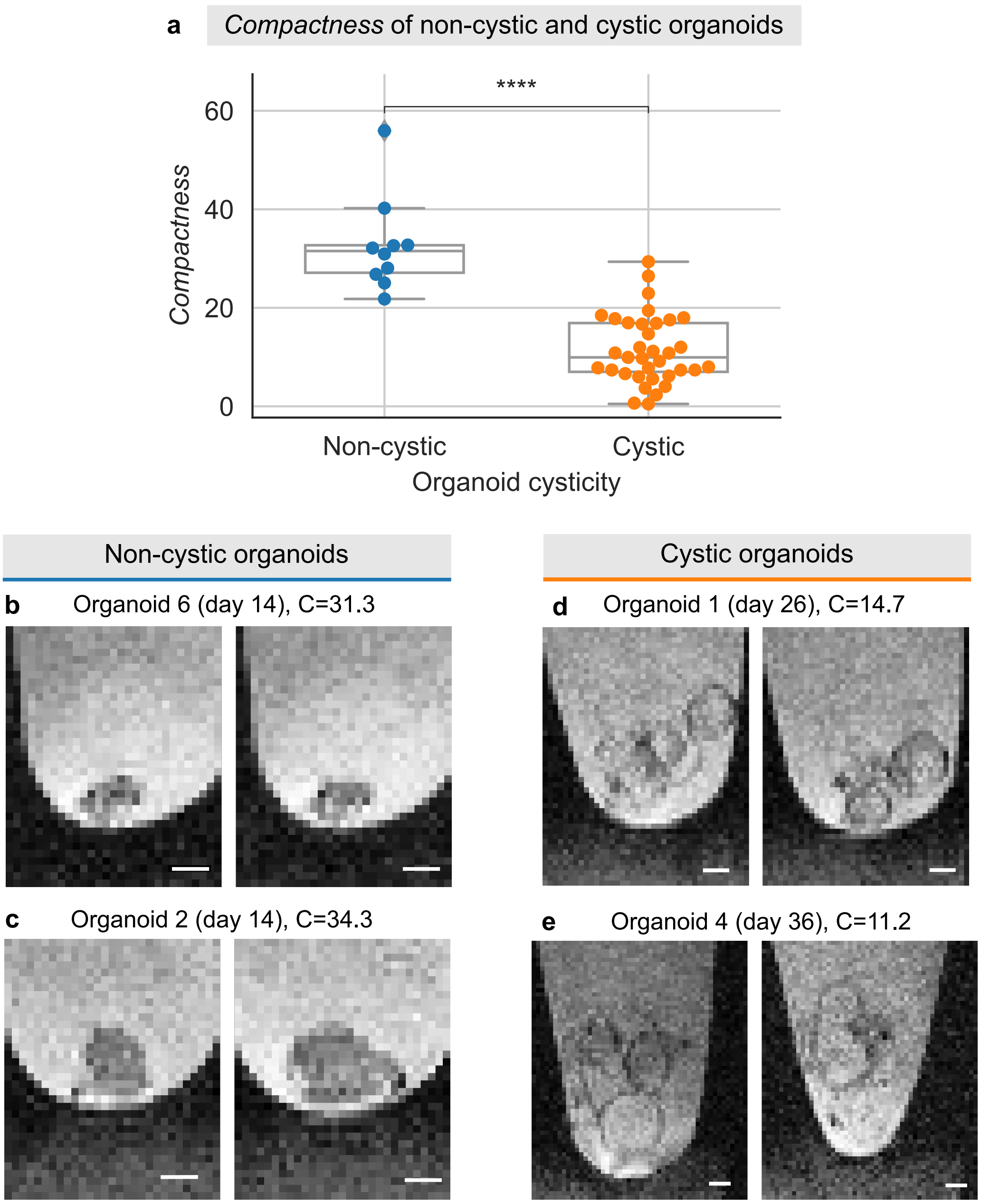


Figure 2.6: Global cysticity classification in MRI. (a), Compactness separates cystic and non-cystic organoids. ****$p<0.0001$ two-sided t-test. (b-e), Selected sagittal planes from two cystic and two non-cystic organoids. Selected sagittal planes (left to right): (b), 59, 60; (c), 41, 45; (d), 58, 61; (e), 36, 53. For better visibility of the organoids, the images were cut to the organoid location. (b-e) *Compactness* is abbreviated with C. Scale bar: 400 μm. Derivative of Fig. 2 in Deininger et al. [178], licensed under CC BY 4.0, with adapted panel (a) and modified layout.

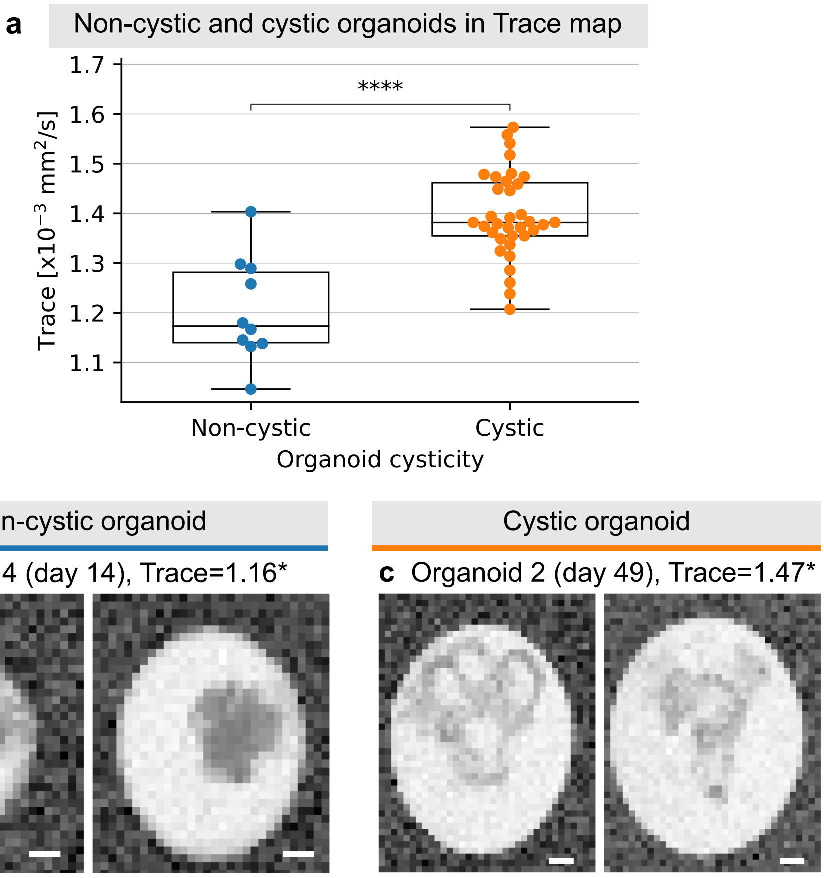


Figure 2.7: Diffusion tensor imaging (Trace map) shows different tissue characteristics of cystic and non-cystic organoids. (a), Trace of non-cystic and cystic organoids. $^{****}p<0.0001$ two-sided t-test and Holm-Šídák correction to adjust for multiple testing of other DTI maps. (b-c), Selected coronal planes from one cystic and one non-cystic organoid; $^{*}[\text{x}10^{-3}$ mm2/s]. Selected coronal planes (left to right): b, 1, 2; c, 5, 6. For better visibility, the images were cut to the Eppendorf tube boundaries. Scale bar: 400 μm. Derivative of Fig. 3 in Deininger et al. [178], licensed under CC BY 4.0, with adapted panel (a) and modified layout.

Table 2.1: Performance for global cysticity classification for T2*-w sequence and selected DTI maps. Mean intensity refers to the mean organoid intensity extracted using ground truth organoid segmentations. For the DTI maps, ground truth segmentations were separately generated, as the segmentations from the T2*-w images could not be transferred due to differing image resolutions.

	Data	Method	ROC AUC
	T2*-w	*Compactness*	0.98
		3D ResNet	0.91
		Mean intensity	0.65
DTI	Trace	Mean intensity	0.91
	1st Eigenvalue		0.93
	2nd Eigenvalue		0.91
	3rd Eigenvalue		0.86
	Fractional Anisotropy		0.63

2.1.4.3. Local Cyst Segmentation

The good performance for global cysticity classification raised the question of whether cysts can be segmented locally, which would provide further insight into cyst distribution and location. For this task, the 3D U-Net achieved a Dice score of 0.63 ± 0.15 (mean ± SD). As shown in Figure 2.8a-b, the Dice scores for individual samples showed a large variation with values ranging from 0.34 to 0.83. The analysis of weak and intermediate model predictions showed discrepancies between model predictions and ground truth, especially for organoids with many small cysts (Figure 2.8c-d). The model performed especially well on images with large, clearly visible, and distinct cysts (Figure 2.8e). Compactness exhibited a strong negative correlation with cystic tissue volume based on ground truth local cyst segmentations (Pearson's $r = -0.74$, Figure A.1), supporting its potential as an estimate for cyst volume.

2.1.4.4. Model Versatility to Other Dataset

Training and testing organoid segmentation and local cyst segmentation on $Data_B$ yielded performance comparable to that on $Data_A$ (Table 2.2). Combining datasets A and B maintained similar performance levels for organoid segmentation but led to improved accuracy in local cyst segmentation compared to training on a single dataset (Table 2.2). Cross-testing models—training on $Data_A$ and testing on $Data_B$, and vice versa—also resulted in comparable performance to training and testing on the same dataset for both tasks (Table 2.2). These results indicate that the models demonstrate robustness, with consistent performance across individual and combined datasets and enhanced accuracy in local cyst segmentation when datasets are merged.

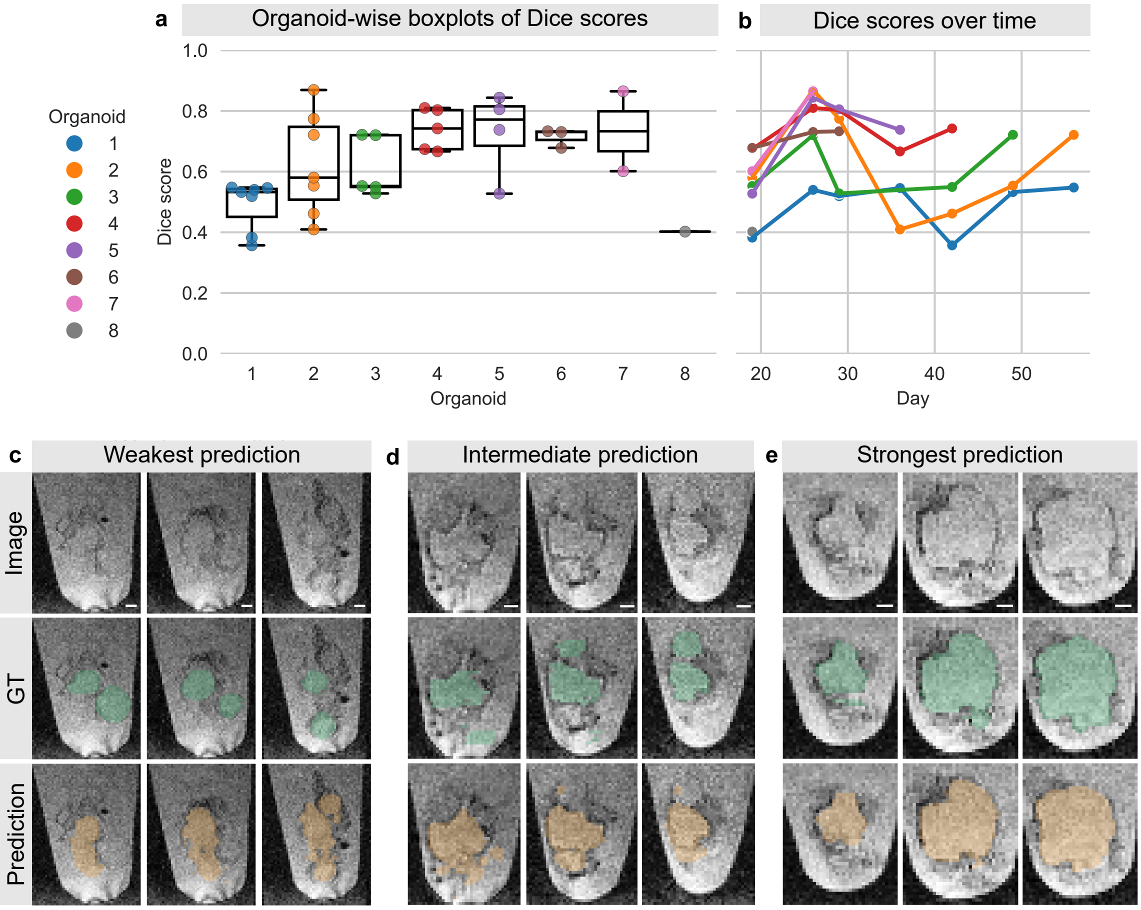


Figure 2.8: 3D local cyst segmentation in MRI. (a-b), Model performance. (c-e), Selected sagittal planes for three organoids. (c), Organoid 1 (day 42): Dice score of 0.34. (d), Organoid 4 (day 36): Dice score of 0.63. (e), Organoid 7 (day 26): Dice score of 0.83. Image: original image, GT: image with ground truth organoid location (green), Prediction: image with predicted organoid location (orange). For better visibility of the organoids and to investigate the detailed differences between ground truth and model prediction, the images were cut to the organoid location. Selected sagittal planes (left to right): (c), 60, 55, 51; (d), 52, 49, 42; (e), 63, 56, 49. Scale bar: 400 μm. Derivative of Fig. 4 in Deininger et al. [178], licensed under CC BY 4.0, with adapted panels (a,b) and modified layout.

Table 2.2: Performance for organoid and local cyst segmentation on different datasets. The Dice score is reported as mean ± SD.

Task	$\text{Data}_{\text{Train}}$	$\text{Data}_{\text{Test}}$	Dice score
Organoid segmentation	A	A	0.92 ± 0.06
	B	B	0.91 ± 0.08
	A ∪ B	A ∪ B	0.91 ± 0.09
	A	B	0.93 ± 0.06
	B	A	0.90 ± 0.06
Local cyst segmentation	A	A	0.63 ± 0.15
	B	B	0.63 ± 0.16
	A ∪ B	A ∪ B	0.68 ± 0.13
	A	B	0.57 ± 0.08
	B	A	0.63 ± 0.15

2.1.4.5. Comparison to Brightfield Microscopy

Quantitatively comparing the performance of automated analysis of different data modalities can give insight into the benefits of one data modality over another. Since MRI acquisition is relatively time-consuming and resource-intensive, it is of particular interest to assess whether significantly simpler and more accessible imaging techniques, such as brightfield microscopy, can yield comparable results for the same tasks. To this end, organoid segmentation and global cysticity classification were also performed on brightfield images based on organoids of the same dataset. Local cyst segmentation, however, was not feasible due to the limited visibility of cysts in the 2D brightfield images (Figure A.5).

For organoid segmentation, Otsu's thresholding method [105] did not produce satisfactory results. Therefore, the SegFormer model [111] was used, implemented in [184], for state-of-the-art 2D segmentation. The model was trained with Adam (learning rate = 1×10^{-4}, weight decay = 1×10^{-1}) for 2,000 iterations with batch size 1 and a weighted sum of binary cross entropy and Dice loss (1:10).

For global cysticity classification, the state-of-the-art neural network ResNet18 [95] implemented in [185] was used for binary classification. The model was trained with Adam (learning rate 1×10^{-6}, weight decay 0) for 30 epochs with batch size 16 and a binary cross entropy loss. On-the-fly image augmentations included random rotation (0-360 degrees), random resized crop (scale $0.3 - 1.0$, ratio 1.0), and ColorJitter (brightness = 0.1, saturation = 0.1, contrast = 0.1) to account for large variations in organoid size and color. Since, in contrast to *Compactness*, the deep-learning-based ResNet18 requires training, organoid-wise LOOCV was used for model evaluation.

Correlating organoid sizes across imaging modalities can increase confidence in the respective measurements and reveal modality-specific biases. The correlation between organoid sizes in MRI and brightfield microscopy was strong when organoids were fully contained within the brightfield imaging frame, but weak when organoids extended beyond the frame, resulting in underestimated sizes in brightfield microscopy (Figure A.4). These protocol-related limitations, along with strategies for their detection and correction, are discussed in detail in Section 2.2.4.2.

Overall, organoid segmentation in MRI performed similarly to brightfield organoid segmentation (Table 2.3, Figure A.2). Global cysticity classification performed slightly better in MRI compared to brightfield images (Table 2.3). However, this is likely also due to the limited dataset available for model training. The MRI experimental setup supports concepts like *Compactness*, which is inspired by the signal-to-signal ratio. In contrast, such concepts do not apply to brightfield imaging due to the heterogeneous microscopy backgrounds and the limited spatial visibility of cysts in 2D images (Figure A.5).

Table 2.3: Comparison of best-performing MRI-based models and brightfield microscopy-based models for organoid segmentation and global cysticity classification. Segmentation performance is reported using the Dice score, while classification performance is reported as the ROC AUC. All metrics are derived from cross-validation, except for Otsu and *Compactness*, which are not machine learning-based and are computed on the full dataset A.

Task	Data	Method	Metric
Organoid segmentation	MRI	3D U-Net	0.92 ± 0.06
	BF	SegFormer	0.91 ± 0.11
	BF	Otsu	0.84 ± 0.18
Global cysticity classification	MRI	*Compactness*	0.98
	BF	ResNet18	0.94 ± 0.08

2.1.5. Discussion

With mean Dice scores of 0.92 and 0.91, the performance of organoid segmentation in 3D MRI images is on par with AI-based organoid segmentation for brightfield imaging. Comparable methods for MRI brain segmentation achieve Dice scores in the range of 0.72 and 0.93 [186–190]. Such a highly reliable automated analysis will represent a powerful tool to compare wild-type organoids with disease models associated with altered growth rates, such as Zika-Virus disease [191] or neurodevelopmental diseases leading to microcephaly [192].

As the first step, reliable organoid segmentation paves the way for comprehensive, non-destructive quality monitoring, including morphological and functional tissue parameters. For MRI, the newly introduced metric *Compactness*, inspired by the concept of signal-to-signal ratio [193, 194], assesses overall cysticity. It successfully separated non-cystic and cystic organoids, closely matching the phenotypical appearance of previously reported non-cystic and cystic organoids (Figures A.5, A.6) [56, 79], at a ROC AUC of 0.98 and outperformed the deep learning-based ResNet18 applied to brightfield imaging (ROC AUC 0.94). In contrast to brightfield imaging, which only provides morphological insight, DTI measures functional tissue parameters. Using DTI, it was shown that cystic organoids have a significantly higher diffusion than non-cystic organoids, most likely reflecting their increased fluid content. This observation aligns with the anatomical differences between the two groups, as fluid-filled cysts facilitate less restricted water diffusion. To further investigate cyst formation, especially using DTI, choroid plexus organoids could be an interesting model [79]. While the study primarily focused on the detection of cystic miss-differentiation, the pipeline could likely also be applied to investigate necrotic core formation in larger, longer-matured cerebral organoids, which could be used as a quality control readout as well.

Successful global cysticity assessment led to the question of whether cysts can be segmented locally to differentiate solid compartments from fluid-filled cavities. While organoid segmentation and global cysticity classification were conducted in brightfield and MRI images for comparison, local cyst segmentation was performed for MRI images only. Due to the 2D nature of brightfield images, reliable annotation of cysts is rather difficult (Figure A.5). Using the 3D MRI images, an annotation of cysts was more feasible, indicating a better assessment of the three-dimensional morphology using MRI images. The 3D U-Net trained for local cyst segmentation reached a mean Dice score of 0.63, which indicates a challenging segmentation task. Other challenging segmentation tasks, such as ischemic stroke lesion segmentation, achieve Dice scores of 0.37 in MRI [195, 196] and 0.54 in CT [195, 197]. MRI acquisitions were performed using a high-end 9.4 Tesla system, of which only a limited number exist worldwide. Despite the relatively high resolution of this advanced system, accurate segmentation of individual cysts, particularly in organoids with many small cystic structures, remains a significant challenge due to inherent limitations in spatial resolution and contrast-to-noise ratio. In such cases, global cysticity classification may thus capture more easily the fluent transition from compact to cystic organoids.

Some limitations need to be taken into consideration. On the one hand, reliable MRI organoid segmentation and global cysticity assessment could be achieved despite the relatively small dataset and heterogeneous organoid morphology, with an equal or better performance compared to state-of-the-art AI-based methods applied to brightfield images. As shown, extending the dataset did not further improve the performance for organoid segmentation. On the other hand, local cyst segmentation slightly improved when extending the dataset (mean Dice score +0.05). However, technical limitations of image acquisition still impede segmentation performance in the case of many small cysts due to uncertainty concerning exact boundary detection for both human annotation and model prediction.

Overall, this work presents the first analysis pipeline for the non-invasive quantification of cerebral organoids based on MRI. It was shown that cerebral organoids can be accurately monitored over time and for quality assessment using state-of-the-art tools for automated image analysis. In comparison to brightfield imaging, MRI gives better insight into 3D cerebral organoid morphology, and DTI provides functional tissue characteristics of cerebral organoids. As a result, this pipeline facilitates iterative comparative analysis of organoids and supports the refinement of cultivation protocols. These findings point out the pipeline's potential for clinical application to larger-scale comparative organoid studies.

2.2. Brightfield Microscopy

> Parts of this section are based on: J. Schröter, L. Deininger, B. Lupse, et al. "A large and diverse brain organoid dataset of 1,400 cross-laboratory images of 64 trackable brain organoids". In: *Scientific Data*, Vol. 11, No. 1 (May 2024). ISSN: 2052-4463. DOI: `10.1038/s41597-024-03330-z`. [198]. J.S. was responsible for brain organoid cultivation and imaging, while L.D. contributed through data annotation, processing, and development of the automated analysis pipeline.

While many analysis pipelines exist for brightfield microscopy of organoids, their sensitivity to imaging artifacts, such as reflections, shadows, and color variations from the culture medium, can compromise segmentation accuracy and hinder cross-laboratory reproducibility. This limits reliable growth tracking and the extraction of morphological features. This section explores how a segmentation and analysis pipeline can be designed for robust and explainable monitoring of brain organoid growth, diversity, and protocol refinement.

As shown in Section 2.1, volumetric analysis via MRI is feasible but experimentally demanding, while brightfield microscopy is the standard approach for routine organoid size assessment due to its simplicity and accessibility. This work presents a robust segmentation and analysis pipeline for brightfield images and compares two classical tools—CellProfiler, a highly versatile platform for biomedical image segmentation and analysis, and OrganoSeg, specifically developed for brightfield segmentation of diverse organoids—with two deep learning-based methods: MOrgAna, tailored for brightfield brain organoid segmentation, and SegFormer, a model not previously applied to organoids. Their performance is evaluated across developmental stages, laboratories, and common artifacts in Section 2.2.3.1 and Section 2.2.3.2. Beyond segmentation, the pipeline's utility is demonstrated for analyzing clone diversity (Section 2.2.3.3) and supporting iterative protocol optimization through optical distortion correction and handling organoid extension beyond imaging boundaries (Section 2.2.4).

2.2.1. Dataset

The dataset consists of over 1,400 images capturing 64 trackable brain organoids, derived from four different clones, imaged at 10 time points over 30 days in two separate laboratories (Figure 2.9a, 2.10). In the context of organoids, the term clones refers to distinct groups originating from different stem cells. The clones included a healthy control (wt2D), two patient-derived clones with TUBA1A- and TUBB2A-associated tubulinopathy, and one patient-derived clone with the neurotransmitter disorder tyrosine hydroxylase (TH) deficiency [199, 200]. For each clone, 16 technical replicates were generated.

To enable analyses of organoid size, growth, and diversity, pixel-level organoid annotations were created (Figure 2.9b). Given the high clone diversity, distinct developmental

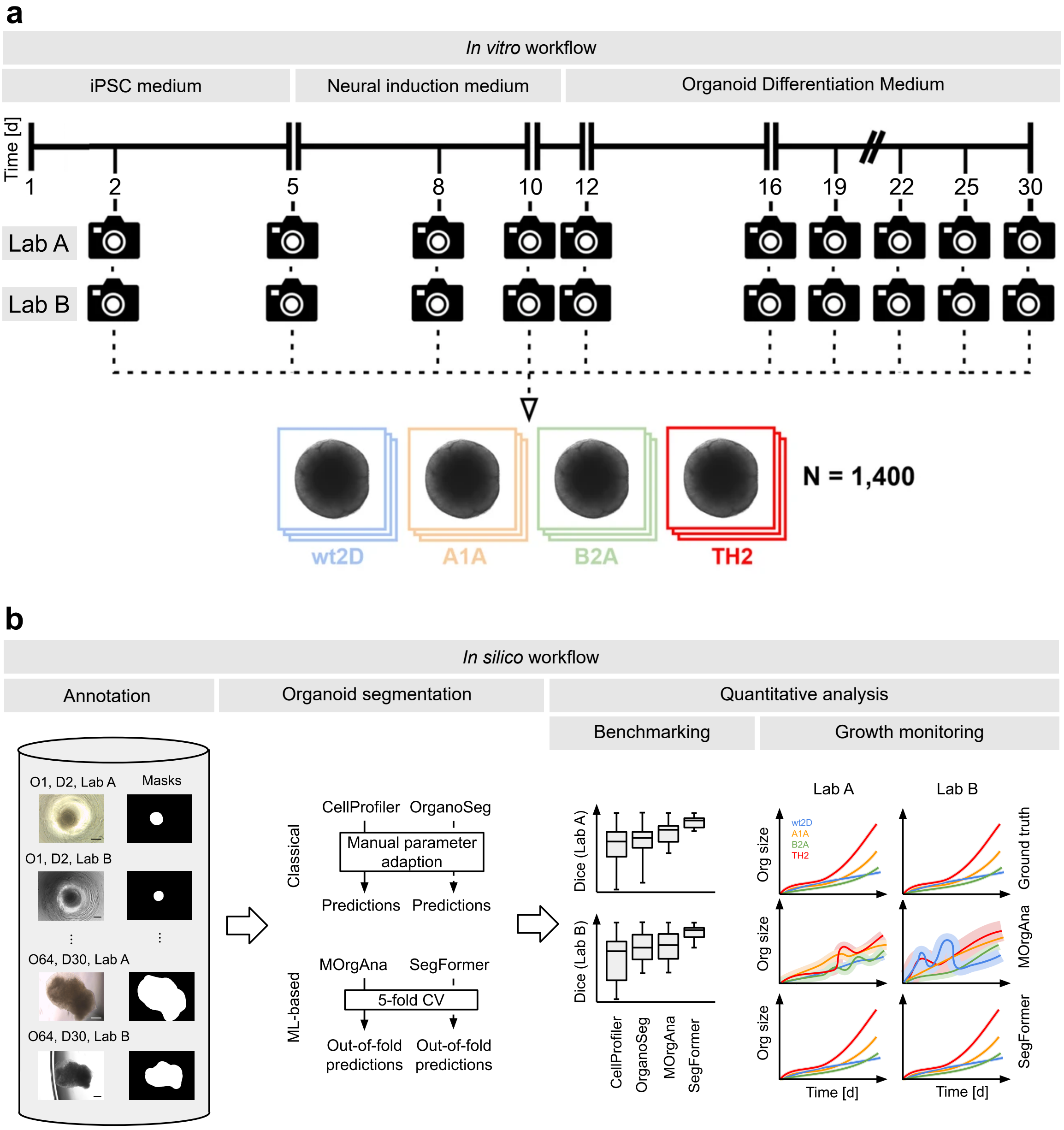


Figure 2.9: Data generation and analyses of organoid analysis pipelines. (a) Replicates of the four clones wt2D, A1A, B2A, and TH2 were cultivated and imaged over 30 days, yielding a dataset of more than 1,400 images. (b) This dataset enables a systematic evaluation of the strengths and limitations of methods for organoid growth monitoring. Scale bar: 500 μm. Derivative from Fig. 1 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with modified layout.

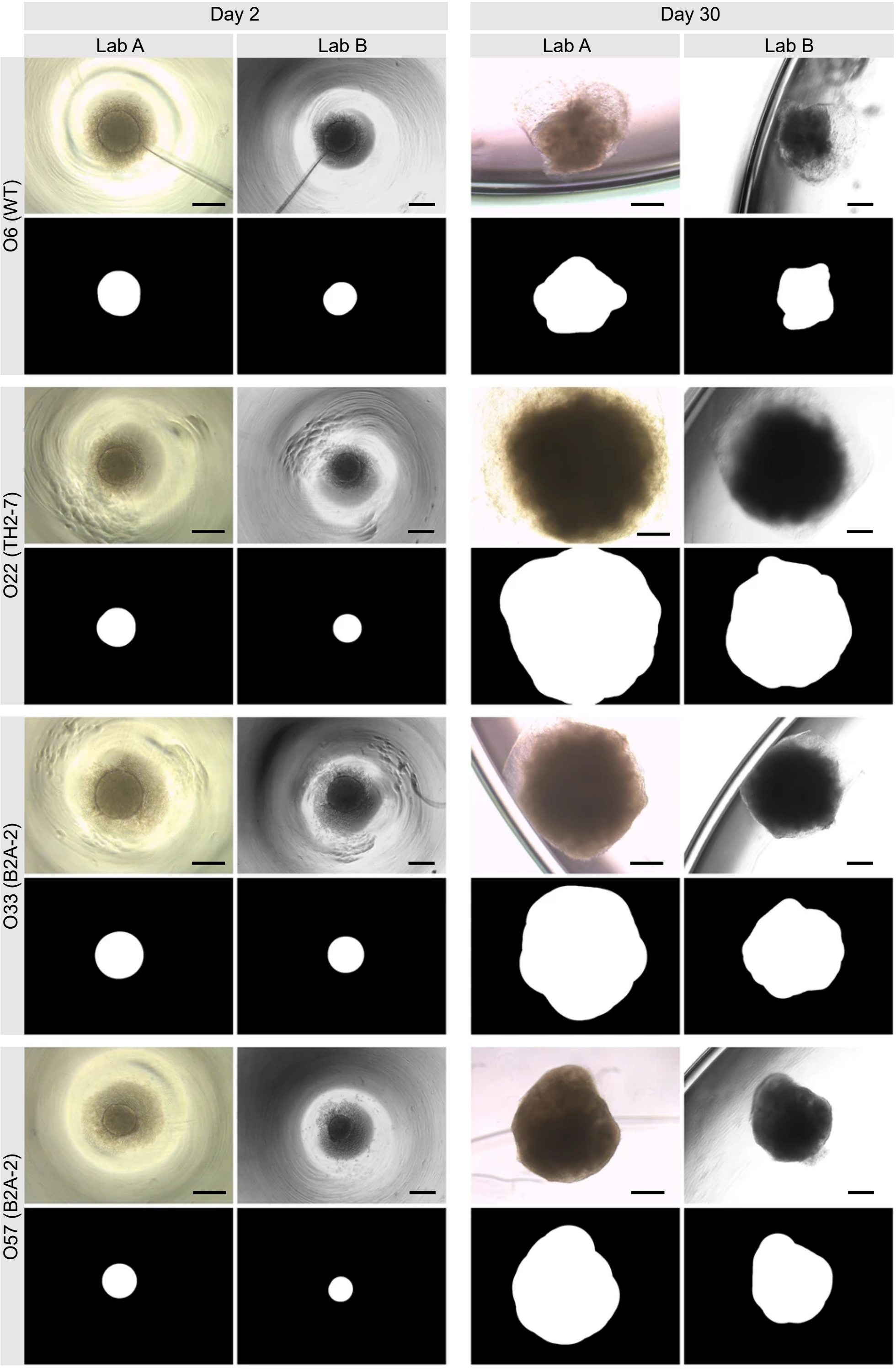


Figure 2.10: Excerpt of the dataset. One representative organoid per clone is shown alongside the corresponding ground truth segmentation for day 2 and day 30, imaged in Lab A and Lab B. The examples illustrate heterogeneous organoid development across time and imaging conditions. On day 2, the dark, textured region surrounding the central circular area indicates cell debris around the embryoid body. By day 30, the translucent structures encircling the organoid correspond to the Matrigel matrix in which the organoids are embedded. Scale bar: 500 μm. Derivative from Fig. 3 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with modified layout.

patterns, cross-laboratory imaging conditions, frequent imaging intervals, and common imaging artifacts—such as light reflections from plate well rims, shadows, and varying colors due to culture medium—this dataset supports detailed evaluation of automated organoid analysis pipelines, providing insights into their specific strengths and limitations.

All images and annotations for organoid segmentation generated for this work are publicly available on Zenodo (`https://zenodo.org/records/10301912`). The code to reproduce the results is publicly available on GitHub (`https://github.com/deiluca/robust_monitoring_organoid_growth/tree/main`).

2.2.2. Methods

2.2.2.1. Organoid Segmentation

CellProfiler

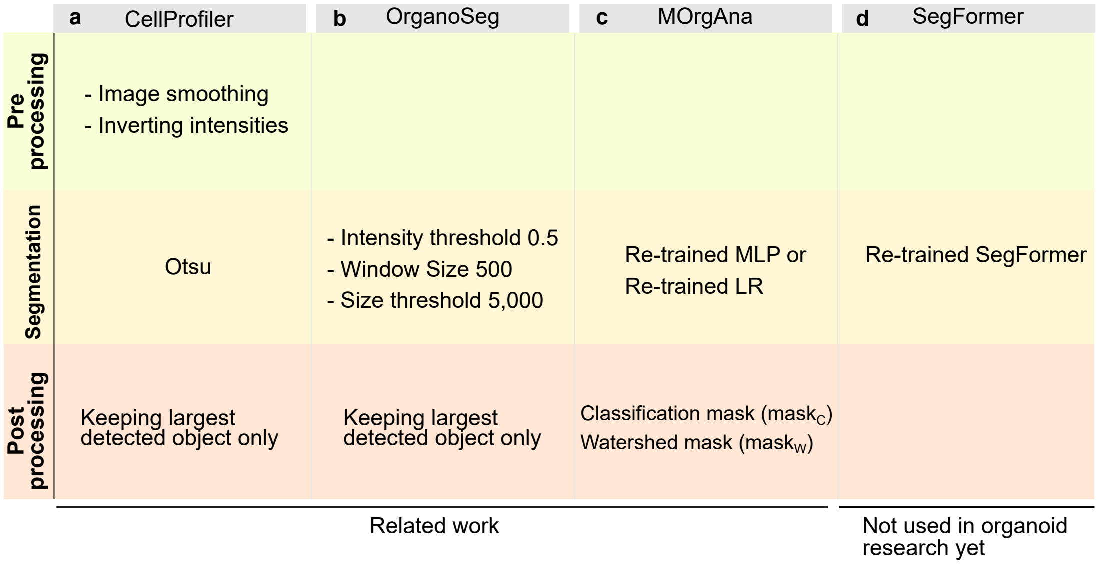
For organoid segmentation, the procedure and parameter settings previously optimized for brain organoids [166] were followed and are outlined here for reference (Figure 2.11a). Initially, image smoothing was performed using morphological opening and closing operations, with a structuring element of 25 pixels in diameter. Next, image intensities were inverted via the ImageMath module. Primary objects were identified using Global Otsu segmentation, applying a two-class thresholding approach. To eliminate debris, analysis was restricted to the largest detected object by applying the MeasureObjectSizeShape method, followed by object filtering with FilterObjects.

OrganoSeg

For brain organoid segmentation, the default OrganoSeg pipeline, previously established in [166], was applied and is summarized here for reference (Figure 2.11b). Segmentation was performed using an Intensity Threshold of 0.5, a Window Size of 500, and a Size Threshold of 5,000. To eliminate debris, any objects smaller than the largest detected object were excluded.

MOrgAna

A primary module within MOrgAna is its organoid segmentation feature, which calculates pixel-wise characteristics and then classifies those pixels for accurate organoid segmentation (Figure 2.11c). The developers implemented two model types for this purpose: a Multilayer Perceptron (MLP) and Logistic Regression (LR). For each model, MOrgAna generates two distinct masks: the *classification mask* ($mask_C$) and the *watershed mask* ($mask_W$). To determine the optimal approach, both mask types were evaluated for the MLP model ($\text{MOrgAna}_{\text{MLP,C}}$ and $\text{MOrgAna}_{\text{MLP,W}}$) and the LR model ($\text{MOrgAna}_{\text{LR,C}}$ and $\text{MOrgAna}_{\text{LR,W}}$), each of which was trained independently. Default parameters were applied to all methods, and training and inference were performed via the MOrgAna graphical user interface. During the evaluation, the organoid border, which MOrgAna separately predicts, was treated as background.



	a CellProfiler	b OrganoSeg	c MOrgAna	d SegFormer
Pre processing	- Image smoothing - Inverting intensities			
Segmentation	Otsu	- Intensity threshold 0.5 - Window Size 500 - Size threshold 5,000	Re-trained MLP or Re-trained LR	Re-trained SegFormer
Post processing	Keeping largest detected object only	Keeping largest detected object only	Classification mask ($mask_C$) Watershed mask ($mask_W$)	

Related work (a–c) | Not used in organoid research yet (d)

Figure 2.11: Overview of preprocessing, postprocessing, and key parameters across organoid segmentation methods. CellProfiler, OrganoSeg, and MOrgAna are established pipelines for organoid analysis, while SegFormer has not yet been applied in this context. CellProfiler and OrganoSeg rely on classical image processing techniques that require manual parameter tuning (a, b), whereas MOrgAna and SegFormer are deep learning-based and were re-trained on the dataset (c, d). Otsu, Otsu's thresholding; MLP, multilayer perceptron; LR, linear regression.

SegFormer

For expedited training, the SegFormer model with the smallest encoder configuration (MiT-B0) was selected (Figure 2.11d). Model training, evaluation, and inference were conducted using the implementation provided in [184]. The model was trained using AdamW, with a learning rate of 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 0.1. Training was performed with a batch size of 2 for a total of 1,000 iterations. A weighted combination of binary cross-entropy and Dice loss (ratio $1 : 10$) was employed to guide the learning process. On-the-fly image augmentation included three steps:

1. Image downscaling to 256×192. This aligns with the MOrgAna default image downscaling.
2. Random horizontal flipping with a probability of 0.5.
3. Z-score normalization (Equation 2.10).

To examine the functionality of the SegFormer model, the encoder's feature maps were analyzed, which represent the features relevant for segmentation in the decoder. To spatially interpret the feature values, organoid segmentations were utilized, and the average feature value relative to the organoid boundary was calculated (Figure 2.12). The first step involved generating organoid segmentations—either from ground truth annotations or, for new data, using predictions from the SegFormer model (Figure 2.12b). Then, a signed Euclidean distance map was created, with positive distances inside the organoid mask that increase further from the boundary, and negative distances outside the mask that decrease with distance from the boundary (Figure 2.12c). For each feature map derived from the SegFormer encoder (Figure 2.12d), feature values were paired with corresponding signed distances. These distances were organized into bins, each representing a specific 1-px-wide range from the boundary, moving inward or outward. The mean feature value was then calculated for each bin, capturing how feature intensity varies as a function of distance from the organoid boundary (Figure 2.12e). For simplicity, the analysis was limited to the first cross-validation (CV) split. Specifically, the model trained on this CV training split was used to extract feature maps on its corresponding CV test split, enabling an investigation into which features the model relies on for prediction. To draw meaningful conclusions, feature values from each feature map were extracted across all test images. For visualization, min-max normalization (Equation 1.10) was applied to each feature map.

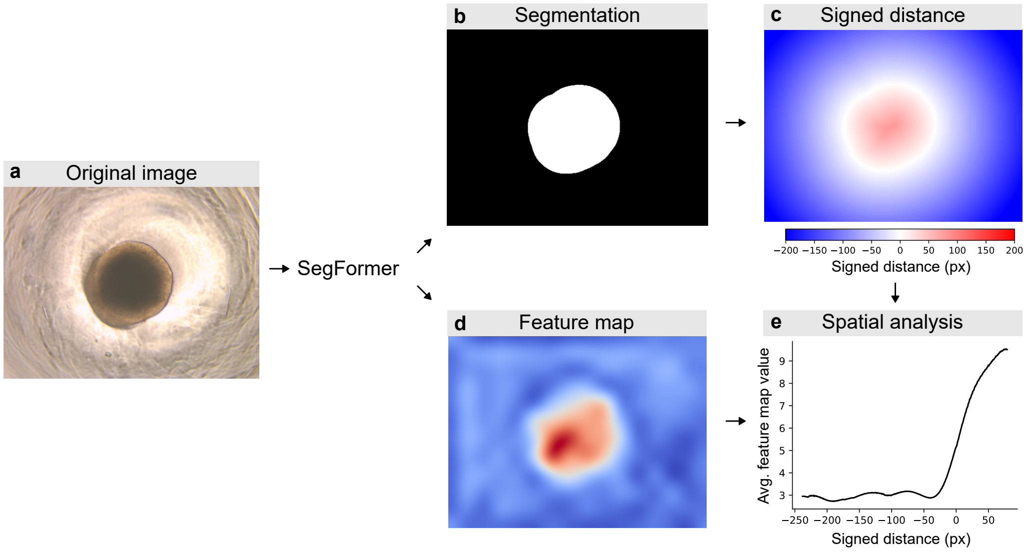


Figure 2.12: Concept of spatially analyzing SegFormer feature values. (a) Input organoid image. (b) Corresponding organoid segmentation, which can be either the ground truth or the segmentation predicted by SegFormer. (c) Signed Euclidean distance map computed from the binary mask in (b), with positive values increasing toward the organoid center and negative values decreasing outward into the background. (d) Feature map extracted from Transformer block 4 of the SegFormer encoder, based on input image (a). (e) Spatially resolved SegFormer feature value as a function of distance from the organoid boundary based on (c) and (d).

Model Evaluation

The Dice score was used to compare the segmentation performance of the models. For unbiased model comparison, 5-fold cross-validation (CV) was used. Since the Dice score provides a sample-level evaluation score, in contrast to e.g., Accuracy, the Dice score distributions of the out-of-fold model predictions were compared. The 5-fold CV splits were organized (1) by organoid to ensure that all images of a particular biological sample were exclusively in either the training or test set, and (2) stratified by clone to reduce model bias toward any specific clone. For SegFormer, the CV training data was randomly partitioned into an 80% training set and a 20% validation set for model selection. In contrast, MOrgAna automatically creates its internal training and validation split from the CV training data.

2.2.2.2. Organoid Growth Monitoring

Organoid size, measured as area in pixels, serves only as a relative metric, as it depends on both the microscopy magnification and the image resolution. To determine the absolute organoid size, the pixel-based organoid size $A_{\text{org}}^{\text{px}^2}$ is converted to micrometers squared $A_{\text{org}}^{\mu\text{m}^2}$ using the scaling relation in Equation 2.7. For this, the corresponding scaling factor from Lab A (Equation 2.8) or Lab B (Equation 2.9) must be used in Equation 2.7.

$$A_{\text{org}}^{\mu\text{m}^2} = A_{\text{org}}^{\text{px}^2} \cdot \left(s_{\mu\text{m}/\text{px}}\right)^2 \tag{2.7}$$

$$s_{\mu\text{m}/\text{px}}^{\text{Lab A}} = \frac{500\,\mu\text{m}}{158\,\text{px}} \tag{2.8}$$

$$s_{\mu\text{m}/\text{px}}^{\text{Lab B}} = \frac{500\,\mu\text{m}}{167\,\text{px}} \tag{2.9}$$

2.2.2.3. Clone Diversity

To quantify morphological differences between clones, 2D organoid features were extracted using PyRadiomics [114]. The features included Elongation, MajorAxisLength, MaximumDiameter, MeshSurface, MinorAxisLength, Perimeter, PerimeterSurfaceRatio, PixelSurface, and Sphericity. To enable visualization of clone diversity, all features were z-score normalized (Equation 2.10) to ensure equal contribution to the subsequent Principal Component Analysis.

$$z = \frac{x - \mu}{\sigma} \tag{2.10}$$

Here, z represents the z-score normalized value, x denotes the original data point, μ is the mean of the dataset, and σ represents the standard deviation of the dataset.

To understand the relationship between the features and the principal components, the loading values of each feature for the principal components were extracted. These loadings indicate the strength of each feature's contribution to the principal components.

2.2.3. Results

2.2.3.1. Organoid Segmentation

Model Performances

For Lab A, $\text{MOrgAna}_{\text{MLP,W}}$ achieved the highest performance across all MOrgAna configurations, surpassing CellProfiler and OrganoSeg (Table 2.4, Figure 2.13a). The lower standard deviation of $\text{MOrgAna}_{\text{MLP,W}}$ indicates more consistent performance across samples compared to the higher variability observed in CellProfiler and OrganoSeg. SegFormer outperformed all models, demonstrating both high accuracy and low variability.

For Lab B, OrganoSeg and $\text{MOrgAna}_{\text{MLP,W}}$ exhibited comparable performances (Table 2.4, Figure 2.13b). The higher standard deviation in OrganoSeg suggests greater variability in performance, whereas $\text{MOrgAna}_{\text{MLP,W}}$ shows more stability. Again, SegFormer demonstrated superior performance in comparison to all models.

The analysis also revealed the limited generalizability of CellProfiler and $\text{MOrgAna}_{\text{MLP,W}}$, which performed better in Lab A compared to Lab B (Table 2.4, Figure 2.13c,d). In general, robust models are expected to yield consistent predictions across different imaging labs. In contrast to SegFormer, which demonstrated high cross-lab consistency, CellProfiler, OrganoSeg, and $\text{MOrgAna}_{\text{MLP,W}}$ exhibited greater variability in performance between labs (Figure 2.13e). Additionally, day-wise segmentation accuracy varied notably in Lab B for CellProfiler, OrganoSeg, and $\text{MOrgAna}_{\text{MLP,W}}$, reflecting substantial day-to-day fluctuations in model performance (Figure 2.13d). Additionally, segmentation quality for CellProfiler and OrganoSeg was particularly poor on Day 2 images (Figure 2.13c,d). In contrast, SegFormer consistently delivered high accuracy across all days and in both labs (Figure 2.13c,d).

Table 2.4: Overview of organoid segmentation performance of all models. The segmentation performance is reported as the Dice score (mean ± SD). Results are based on 5-fold cross-validation.

Model	Lab A	Lab B	Overall
CellProfiler	0.77 ± 0.30	0.53 ± 0.39	0.65 ± 0.36
OrganoSeg	0.75 ± 0.36	0.77 ± 0.34	0.76 ± 0.35
$\text{MOrgAna}_{\text{LR,C}}$	0.72 ± 0.13	0.59 ± 0.17	0.65 ± 0.17
$\text{MOrgAna}_{\text{LR,W}}$	0.86 ± 0.17	0.73 ± 0.27	0.79 ± 0.23
$\text{MOrgAna}_{\text{MLP,C}}$	0.84 ± 0.16	0.71 ± 0.24	0.78 ± 0.22
$\text{MOrgAna}_{\text{MLP,W}}$	0.88 ± 0.15	0.76 ± 0.24	0.82 ± 0.21
SegFormer	0.96 ± 0.05	0.96 ± 0.03	0.96 ± 0.04

When SegFormer was trained and tested using 5-fold cross-validation on a combined dataset from both Lab A and Lab B, it achieved a Dice score of 0.96 ± 0.05 (Table 2.5). Training exclusively on Lab A's dataset and testing on Lab B's dataset resulted in a slightly lower Dice score of 0.94 ± 0.08, suggesting a minor reduction in performance when generalizing to unseen data (Table 2.5). Similarly, training on Lab B's dataset and testing

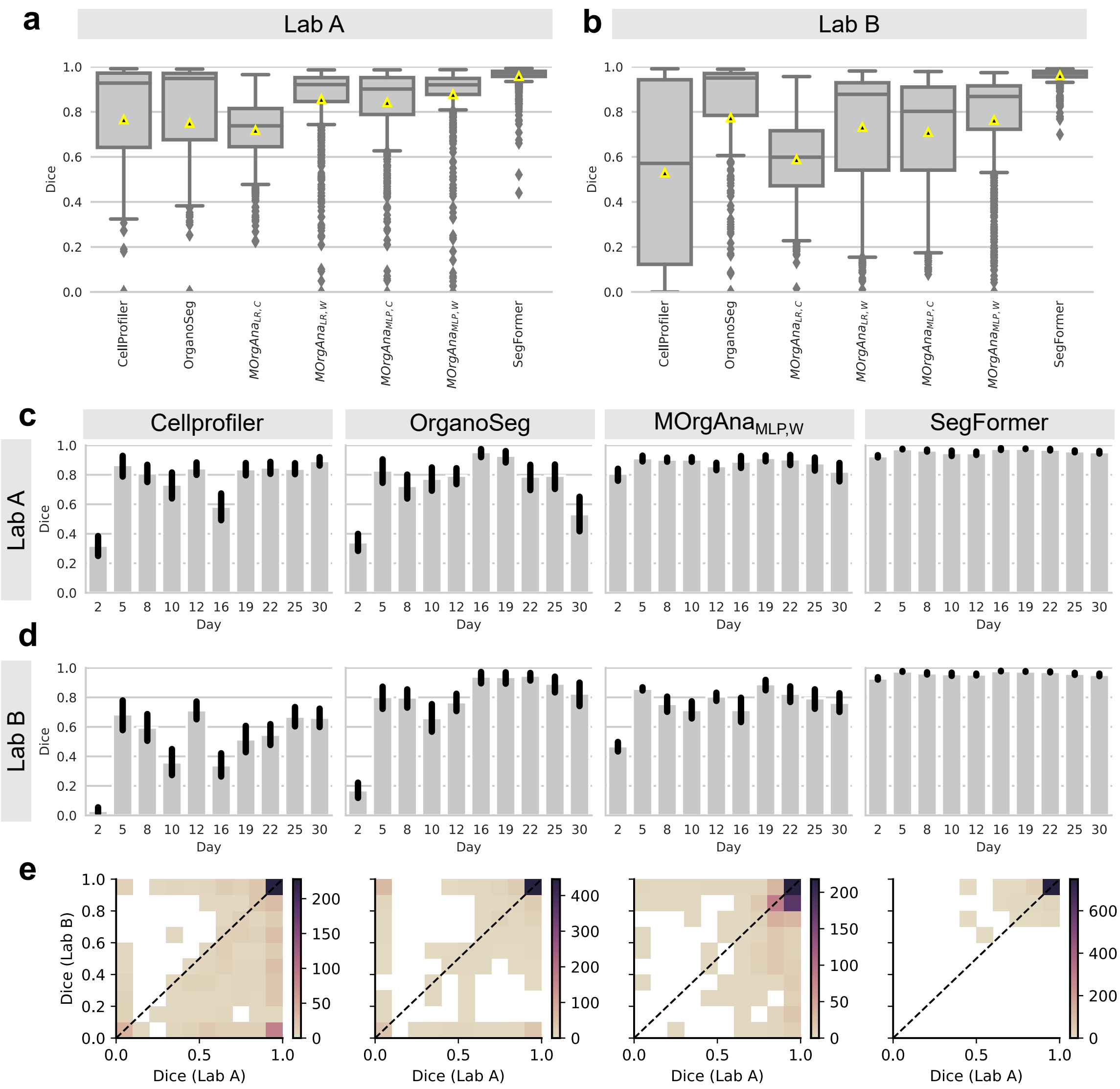


Figure 2.13: Segmentation performance of different organoid analysis pipelines. (a,b) Dice score distribution of all models by imaging lab. Yellow triangle: mean Dice score. (c,d) Day-wise Dice scores for CellProfiler, OrganoSeg, best MOrgAna model, and the SegFormer for both imaging labs. (d) has the same model order as (c). (e) 2D histogram showing the correlation of Dice scores between Lab A and Lab B across models. High values along the diagonal and limited spread in off-diagonal areas indicate strong agreement in predictions between labs. Model order corresponds to (c); bin width is set to 0.1. (c,d) Error bars represent 95% confidence intervals across images acquired on the same day and at the same imaging lab (n = 64 for all days, except Day 12 with n = 128 due to imaging before and after well transfer). Adapted from Fig. 4 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with added panel (e) and modified layout.

on Lab A's dataset yielded a Dice score of 0.94 ± 0.07, reflecting consistent performance across different datasets (Table 2.5). These findings underscore SegFormer's ability to maintain high segmentation accuracy and generalizability, particularly when trained on diverse datasets.

Table 2.5: Cross-testing SegFormer for organoid segmentation. #1: results obtained using 5-fold cross-validation (CV) across the entire dataset, incorporating images from both Lab A and Lab B. #2: utilized the same 5-fold CV splits as #1. However, training and validation were conducted exclusively on images from Lab A, with testing performed solely on images from Lab B. #3: employed the same 5-fold CV splits as #1. Training and validation were restricted to images from Lab B, while testing was conducted exclusively on images from Lab A. The Dice score is reported as mean ± SD.

Exp	$\text{Lab}_{\text{Train}}$	Lab_{Test}	Dice score
#1	$A \cup B$	$A \cup B$	0.96 ± 0.04
#2	A	B	0.94 ± 0.08
#3	B	A	0.94 ± 0.07

Model Robustness to Diverse Phenotypes and Imaging Artifacts

The inclusion of diverse organoid phenotypes and common imaging artifacts in the dataset provides a basis for evaluating the robustness of organoid analysis pipelines. An initial examination shows the robustness of these pipelines when analyzing day-2 matrigel-surrounded organoids (Figure 2.14). For one of these organoids in Lab B, CellProfiler segmented only the background (Figure 2.14a). Moreover, CellProfiler, OrganoSeg, and $\text{MOrgAna}_{\text{MLP,W}}$ in both labs often misclassified the surrounding matrigel as part of the organoid, while SegFormer correctly identified the organoid boundary (Figure 2.14a).

CellProfiler and $\text{MOrgAna}_{\text{MLP,W}}$ occasionally misclassified dark background as an organoid, notably for organoids imaged on days 2, 8, 10, 16, and 30 (CellProfiler, Figure 2.14a-e) and on days 10 and 30 ($\text{MOrgAna}_{\text{MLP,W}}$, Figure 2.14c,e). On day 30 in Lab A, SegFormer incorrectly segmented two organoids instead of one (Figure 2.14e).

SegFormer Explainability

The SegFormer architecture consists of an encoder and a decoder. The encoder is composed of four transformer blocks that process image patches in multiple stages, allowing the model to capture information at different scales (Figure 2.15a). Analyzing the feature maps of individual organoids revealed distinct roles for each block. Block 1 captured fine details and edges, outlining the organoid structures broadly (Figure 2.15b). Blocks 2 and 3 exhibited more focused patterns, emphasizing the organoid boundary and isolating it from the background, thus enhancing its contour (Figure 2.15b). Block 4 was most concentrated on the organoid's central region (Figure 2.15b).

Systematic analysis across all test images in a cross-validation split supported these observations (Figure 2.15d). Feature maps from Blocks 1 and 2 displayed high focus at the

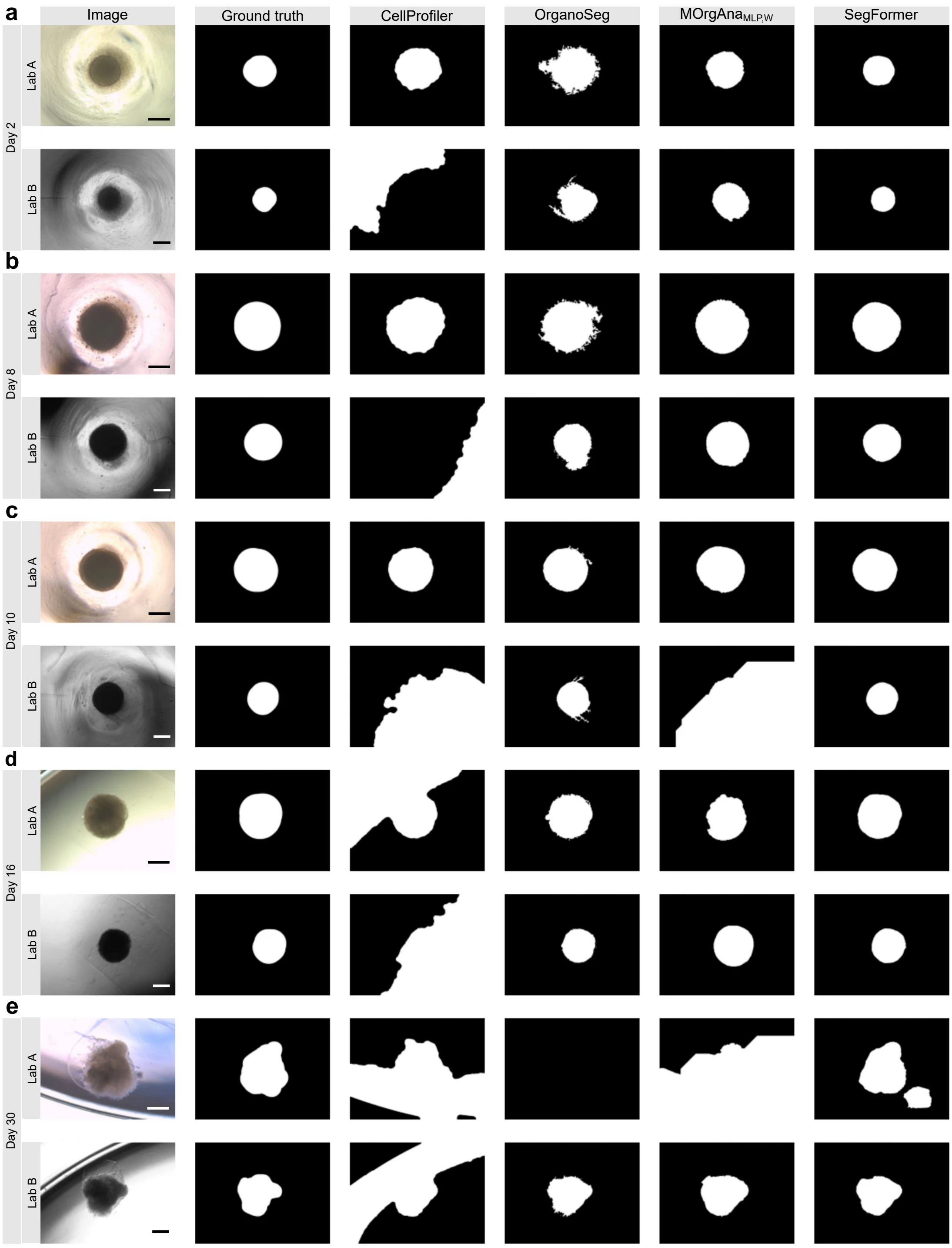


Figure 2.14: Robustness of organoid analysis pipeline to heterogeneous organoid phenotypes and different imaging features. Shown are images of organoid 3 for days 2 (a), 8 (b), 10 (c), 16 (d), and 30 (e), for both imaging labs (first column). Columns 2–6 display the ground truth (GT) segmentation, followed by predictions from CellProfiler, OrganoSeg, $\text{MOrgAna}_{\text{MLP,W}}$, and SegFormer, respectively. On day 2, the dark textured region around the centered circular region represents cell debris around the embryoid body. On day 30, the translucent circumferential structures are the Matrigel matrix in which the organoids are embedded. Scale bar: 500 μm. Adapted from Fig. 5 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with modified layout.

organoid boundary, with clear peak values. Feature map 3 also peaked at the boundary, though with a broader distribution and a less distinct peak. Feature map 4, however, showed a shift in peak values approximately 50 pixels from the boundary, indicating a focus on the organoid core.

This analysis illustrates how the SegFormer model progressively refines its attention on the organoid, starting from fine-grained details of the organoid border in early blocks to core-focused features in later blocks. This hierarchical approach enables accurate segmentation by integrating features at varying detail levels.

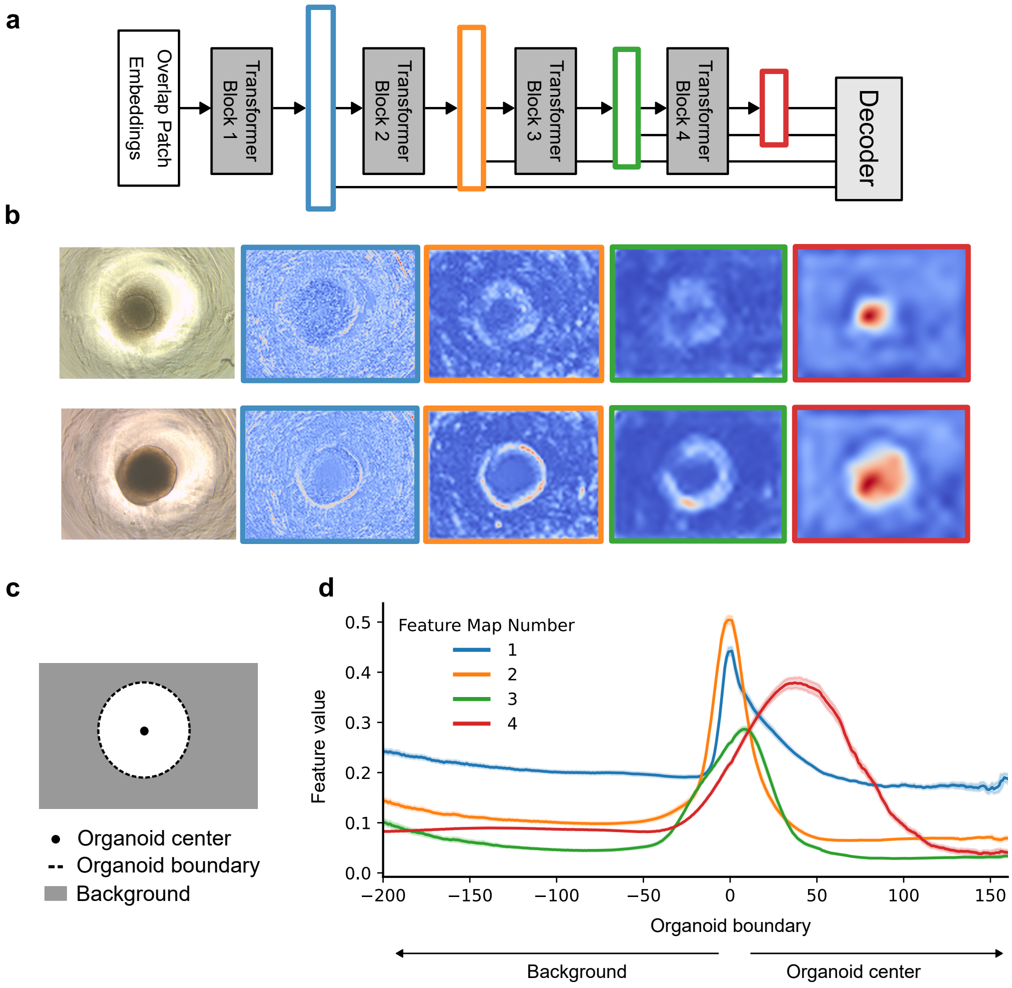


Figure 2.15: Analysis of SegFormer encoder feature values. (a) The SegFormer encoder architecture shows the four transformer blocks, their output stages, and input to the decoder. (b) Example feature maps from each of the four transformer blocks for two representative organoids. (c) Simplified schematic of the organoid structure, depicting the center, boundary, and surrounding background to provide spatial reference for panel (d). (d) Line plot of average feature values from each transformer block's feature map, shown as a function of distance (px) from the organoid boundary. Error bars represent 95% confidence intervals of feature values across test images from one cross-validation split ($n = 286$). (a) adapted from [111], licensed under CC BY 4.0.

2.2.3.2. Organoid Growth Monitoring

Among the clones, TH2-7 demonstrated the most rapid growth trajectory, followed by A1A-1 (Figure 2.16). Clone B2A-2, while initially growing slower than wt2D, eventually reached a comparable size by the end of the observation period. For effective organoid growth monitoring, an ideal model should accurately capture these unique growth patterns. The marked decrease in measured organoid size between days 5 and 8 is attributable to the transfer of organoids from V-bottom 96-well plates to flat-bottom 24-well plates. V-bottom plates tend to optically enlarge the organoids, creating an apparent drop in size when they are transferred.

The SegFormer model stood out as the most accurate, closely aligning with the ground truth measurements for organoid growth in both imaging labs (Figure 2.16). In contrast, OrganoSeg in Lab B provided only a rough approximation of the actual growth pattern, while in Lab A, CellProfiler, OrganoSeg, and $\text{MOrgAna}_{\text{MLP,W}}$ showed significant deviations from the ground truth and did not consistently replicate the observed organoid growth trajectories. CellProfiler, in particular, exhibited considerable fluctuations, especially in Lab B, where the growth trends varied widely. Similarly, $\text{MOrgAna}_{\text{MLP,W}}$ showed notable variability in Lab B, with visible fluctuations that detract from reliable trend representation.

While Lab A generally showed more stable growth trends across the different methods, Lab B displayed greater variability, with CellProfiler and $\text{MOrgAna}_{\text{MLP,W}}$ being the most inconsistent (Figure 2.16b). In comparison, the SegFormer model demonstrated the most reliable and consistent measurements relative to the ground truth across both labs, making it the preferred method for accurate organoid growth monitoring.

Beyond visual inspection, the ground truth organoid annotations allow for the quantitative assessment of each model's accuracy by calculating the maximum day-wise deviation from the actual organoid size (Table A.2). The SegFormer consistently outperformed the other models, achieving the lowest deviations. Specifically, SegFormer demonstrated superior accuracy on 8 out of 10 days in Lab A and across all days in Lab B, with a maximum day-wise deviation of ±7% from the ground truth organoid size. In contrast, the other models—CellProfiler, OrganoSeg, and $\text{MOrgAna}_{\text{MLP,W}}$ —exhibited considerably higher deviations from the ground truth. CellProfiler showed a maximum day-wise deviation of 1768%, while OrganoSeg and $\text{MOrgAna}_{\text{MLP,W}}$ deviated by up to 303% and 351%, respectively (Table A.2).

Overall, these findings underscore the importance of selecting a robust analysis model, as the chosen method can significantly impact the accuracy and consistency of organoid growth assessment, with SegFormer demonstrating a substantial performance advantage over the other models and proving to be the most reliable and precise model for tracking organoid growth.

2.2.3.3. Clone Diversity

To explore clone diversity, a principal component analysis (PCA) was performed across all samples and both imaging labs, using 2D imaging features that capture organoid shape

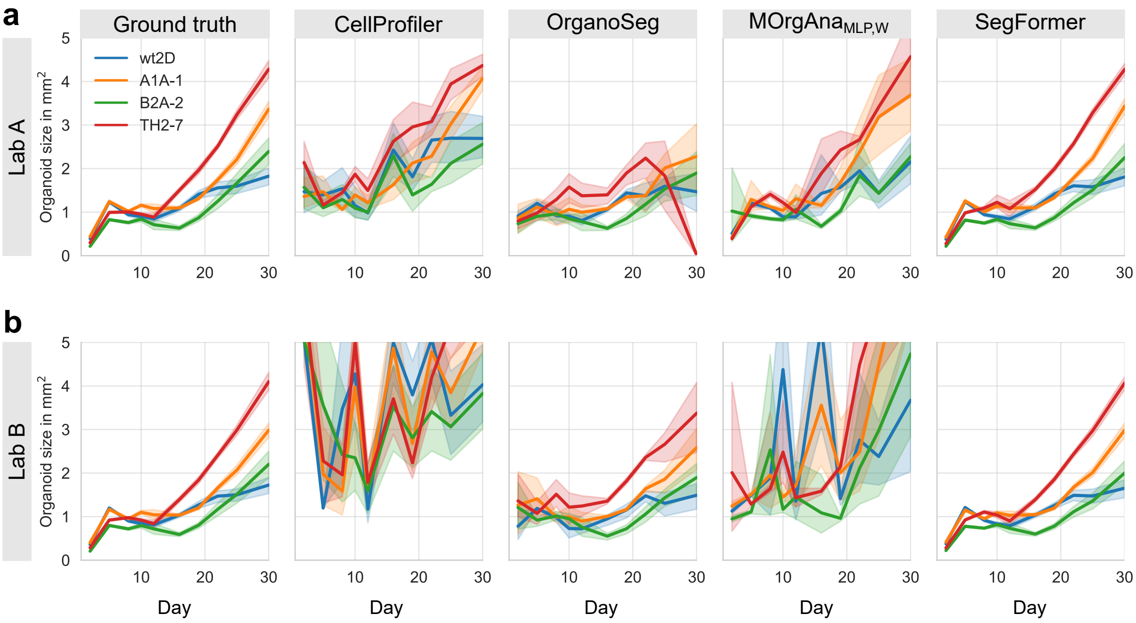


Figure 2.16: Organoid growth based on ground truth and model predictions. Absolute organoid sizes for Lab A (a) and Lab B (b). Columns represent ground truth organoid size (column 1), organoid size based on CellProfiler segmentations (column 2), organoid size from OrganoSeg segmentations (column 3), organoid size from MOrgAna$_{MLP,W}$ segmentations (column 4), and organoid size from SegFormer segmentations (column 5). Adapted from Fig. 6 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with modified layout.

and morphology (Figure 2.17a). The first two principal components accounted for 93.3% of the total variance. Both the combined PCA across all days and the individual day-wise PCAs indicated distinct morphological differences among the clones (Figure 2.17b,d). Notably, clones A1A-1, TH2-7, and B2A-2 consistently displayed clear separation from others on most days (specifically days 2, 5, 12, 16, 19, 22, and 25), highlighting their unique morphological characteristics. Clone WT2D also showed similarity to A1A-1 on days 2, 5, 8, 16, 19, 22, and 25, suggesting overlapping morphological traits at these time points.

To better understand the relationship between the features and the principal components, the loading values were analyzed, which indicate the contribution of each feature to the principal components. This analysis revealed that features dominated by size, such as MajorAxisLength, MinorAxisLength, Perimeter, MeshSurface, PixelSurface, and MaximumDiameter, drove the variance captured by PC1 (Figure 2.17c). In contrast, PC2 was primarily influenced by shape-related features, specifically Sphericity and Elongation.

2.2.4. Protocol Optimization

2.2.4.1. Optical Distortion

During cultivation, organoids grow, and common protocols describe transferring organoids at specific time points into larger wells that can accommodate their increased size or embedding them in Matrigel [47, 56, 198]. However, different wells and Matrigel embedding often have distinct optical properties that may cause artifacts, such as apparent enlargement, complicating organoid growth monitoring.

Automatically and accurately monitoring organoid growth can reveal such optical differences before and after the organoid transfer (Figure 2.18). In the dataset, organoids were transferred from 96-well plates to larger 12-well plates on day 12, where they were embedded in Matrigel. Organoids appeared larger before transfer (Figure 2.18a-d), which introduced challenges in growth monitoring and raised the question of how to standardize organoid size measurements. Importantly, this apparent enlargement affected all clones equally, with no bias toward a specific clone (Figure 2.18c,d), indicating that relative sizes between clones remained comparable while only the absolute sizes were impacted. However, the organoid size before embedding showed a higher standard deviation compared to after embedding, for Lab A (0.22 vs. 0.14) and Lab B (0.21 vs. 0.13). This can make it harder to detect subtle differences between clones, especially if the variability within groups contributes biologically valuable information. With lower variability, the data may appear more uniform, which could lead to the impression that clones are more similar than they truly are.

Three different approaches were considered for analyzing organoid size on day 12:

1. Using *before transfer* measurements yields: (i) growth from day 10 to day 12 for all clones and (ii) shrinkage from day 12 to day 16 in clones B2A-2 and A1A-1, while clones wt2D and TH2-7 showed growth (Figure 2.18e).

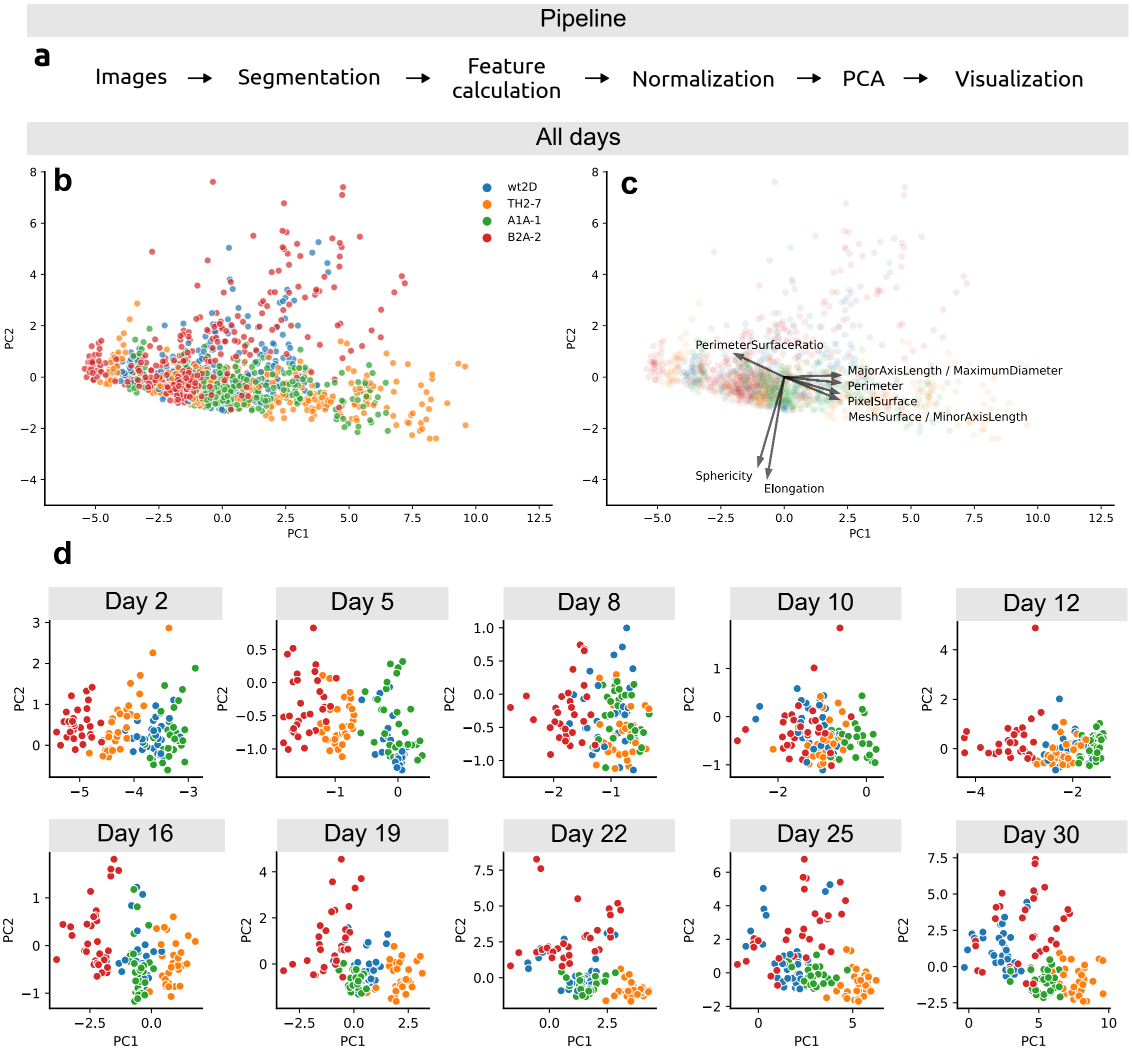


Figure 2.17: Segmentation-based pipeline using PCA for analyzing clone diversity. (a) The analysis pipeline includes image segmentation, extraction of nine 2D imaging features from PyRadiomics characterizing organoid shape and morphology, z-score normalization of features, and dimensionality reduction via principal component analysis (PCA). (b, c) Displayed are the first two principal components from PCA conducted across all samples from both imaging laboratories, using ground truth organoid segmentations. (d) day-wise subsets of (b). (c) Additionally includes loading vectors for each imaging feature, where certain features are combined into a single arrow due to spatial limitations and high similarity in loading values. Adapted from Fig. 7 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with new panels (a-c) and a modified layout in (d).

2. Using *after transfer* measurements results in: (i) shrinkage from day 10 to day 12 for all clones, followed by (ii) growth for all clones from day 12 to day 16 (Figure 2.18f).
3. Averaging *before and after transfer* sizes gives: (i) a slight shrinkage from day 10 to day 12 for all clones, (ii) shrinkage for clones B2A-2 and A1A-1 with growth for clones wt2D and TH2-7 from day 12 to day 16 (Figure 2.18g), and (iii) to a relatively smooth transition from day 8 over day 12 to day 16.

Altogether, averaging before and after transfer seems to be a good tradeoff when aiming for a smooth transition. However, cultivating organoids in the same wells and continuously with/without embedding throughout the observation period would ensure more consistent measurements. If well transfers are unavoidable, measuring organoids both before and after the transfer can help account for well- or embedding-specific optical effects. Additionally, it is advisable to avoid aligning well transfers with critical measurement points to prevent potential biases in size analysis. If transfers are necessary, they should be performed simultaneously for all clones to ensure consistency across experimental conditions. Future studies should investigate clone-specific optical distortions in other well and embedding types, as such effects could introduce substantial biases in growth analysis.

For future investigations, each factor potentially influencing optical distortion should be analyzed independently. Let W1 and W2 represent wells 1 and 2, respectively, and E1 and E2 denote *without embedding* and *with embedding*. In the current study, W1E1 (well 1 without embedding) was compared to W2E2 (well 2 with embedding), noted as W1E1-W2E2. To delineate the optical properties of each factor, all combinations should be examined:

- W1E1-W1E2: To assess the optical properties of different embeddings within well 1.
- W2E1-W2E2: To evaluate the optical properties of different embeddings within well 2.
- W1E1-W2E1: To investigate the optical properties of the two wells, keeping the embedding fixed at E1.
- W1E2-W2E2: To compare the optical properties of the two wells, keeping the embedding fixed at E2.

This systematic approach ensures that the influence of each factor can be clearly understood.

2.2.4.2. Organoid Extension Beyond Imaging Borders

To continuously monitor organoids using brightfield microscopy and automated segmentation, organoids should always be entirely visible for each recording. In practice, however, organoids can grow unexpectedly large, exceeding the limits of the selected microscope or magnification. Standard protocols typically lack intrinsic quality control mechanisms to detect such extension beyond imaging boundaries.

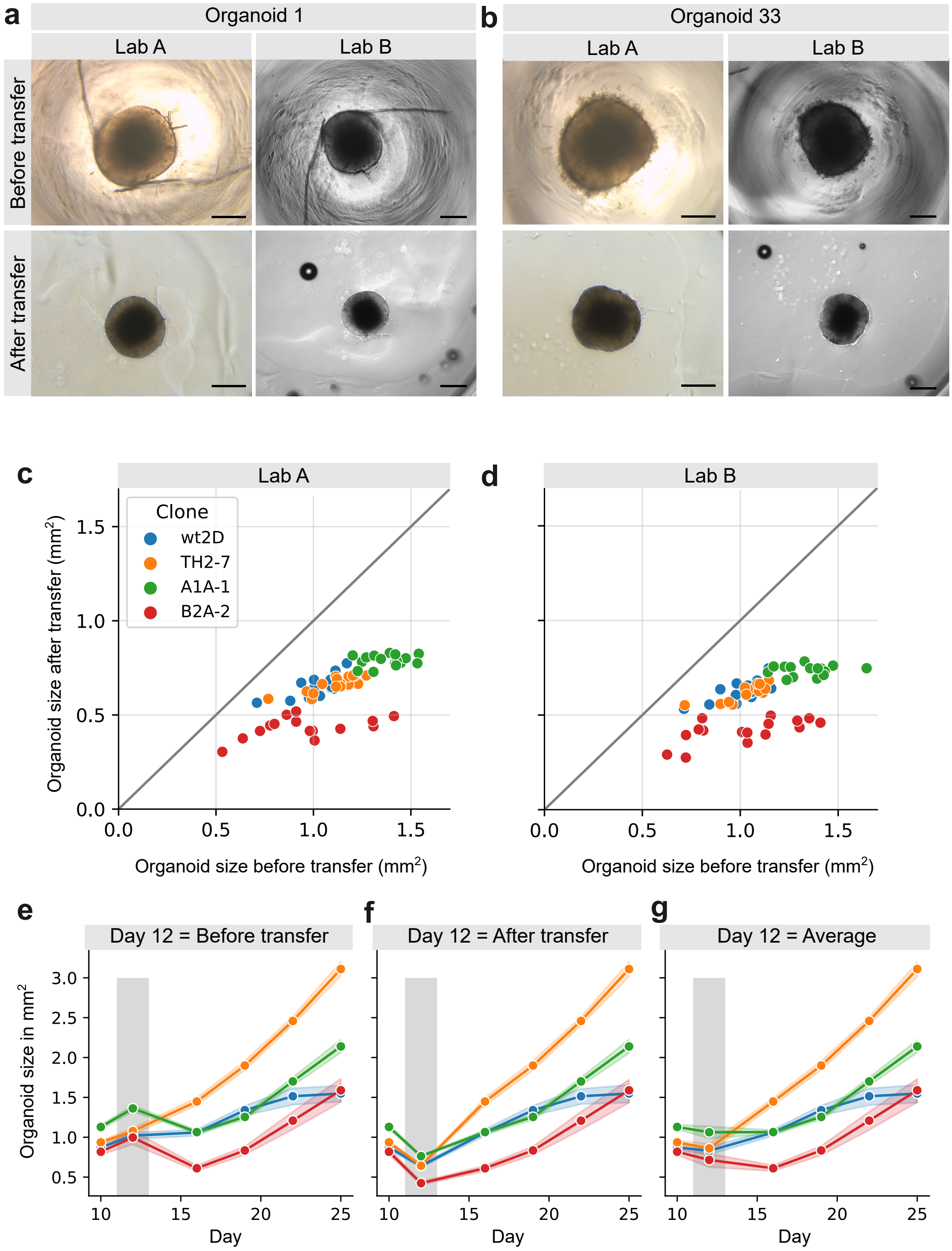


Figure 2.18: Investigation and handling of well-transfer and embedding-induced optical distortion. (a-b) Images of organoids from Lab A and Lab B before and after transfer on day 12, shown for organoid 1 (a) and organoid 3 (b). (c-d) Comparison of organoid sizes before and after transfer on day 12 for Lab A (c) and Lab B (d). (e-g) Development of organoid size using three measurement strategies for day 12: images before transfer (e), images after transfer (f), and averaging the sizes from before and after transfer (g). Scale bar: 500 μm. Adapted from Fig. S1 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with added panels.

Wildtype brain organoids were monitored on days 13, 15, 19, 26, and 29 of organoid cultivation. A period of rapid organoid growth was observed, particularly between days 13 and 19, followed by a plateau in growth after day 19 (Figure 2.19a). Further examination of organoids and their corresponding segmentation masks revealed that organoids were extending beyond the imaging borders (Figure 2.19c). This observation suggests that the apparent growth plateau may not reflect a biological phenomenon but rather an artifact introduced by the imaging protocol. Pixels classified as organoid at the image edge serve as indicators of organoid growth extending beyond the imaging field. To automate the detection of such growth, the ratio of segmentation mask pixels at the image edge classified as organoid can be calculated:

$$\text{Edge ratio} = \frac{\text{Edge sum}}{\#\text{ Edge pixels}} \tag{2.11}$$

$$\begin{aligned} \text{Edge sum} = &\sum_{j=1}^{n} \big(M(1,j) + M(m,j)\big) \\ &+ \sum_{i=1}^{m} \big(M(i,1) + M(i,n)\big) \\ &- M(1,1) - M(1,n) - M(m,1) - M(m,n) \end{aligned} \tag{2.12}$$

$$\#\text{ Edge pixels} = 2n + 2m - 4 \tag{2.13}$$

Where $M(i,j)$ is the binary mask value at row i and column j, with $M \in \{0,1\}^{m\times n}$ with values 0 (background) and 1 (organoid), where m is the number of rows and n is the number of columns. Equation 2.12 sums the pixel values in the top and bottom rows (first line), sums the pixel values in the left and right columns (second line), and subtracts corner pixels to avoid double-counting (third line).

The Edge ratio provides a means to evaluate the overall quality of the dataset. In the dataset, all organoids on days 13 and 15 exhibited 0% Edge ratio, followed by an increase on subsequent days, reaching up to 100% on day 29 (Figure 2.19b). For organoid 6, both size and the Edge ratio decreased from day 26 to day 29. This reduction may be attributed to a folding of the organoid along the z-axis, causing an apparent decrease in size, or a displacement of the organoid outside the imaging area (Figure 2.19a-c).

Assuming that an organoid is always a connected cell compartment, a 0% Edge ratio indicates, under the assumption of accurate segmentation and that the organoid is within the imaging field, that the entire organoid has been imaged and segmented. One can exclude all organoids with >0% Edge ratio to keep only full-imaged organoids (Figure 2.19d). Using a higher Edge ratio can accommodate more organoids, but at a higher uncertainty for accurate organoid size measurements (Figure 2.19e).

Organoid size strongly correlated with the Edge ratio ($r = 0.9, p = 2.3 \times 10^{-34}$). This observation aligns with the expectation that as the organoid grows larger, it becomes more likely to touch the imaging border. However, it is important to note that a higher Edge ratio does not necessarily imply that a larger proportion of the organoid is located outside

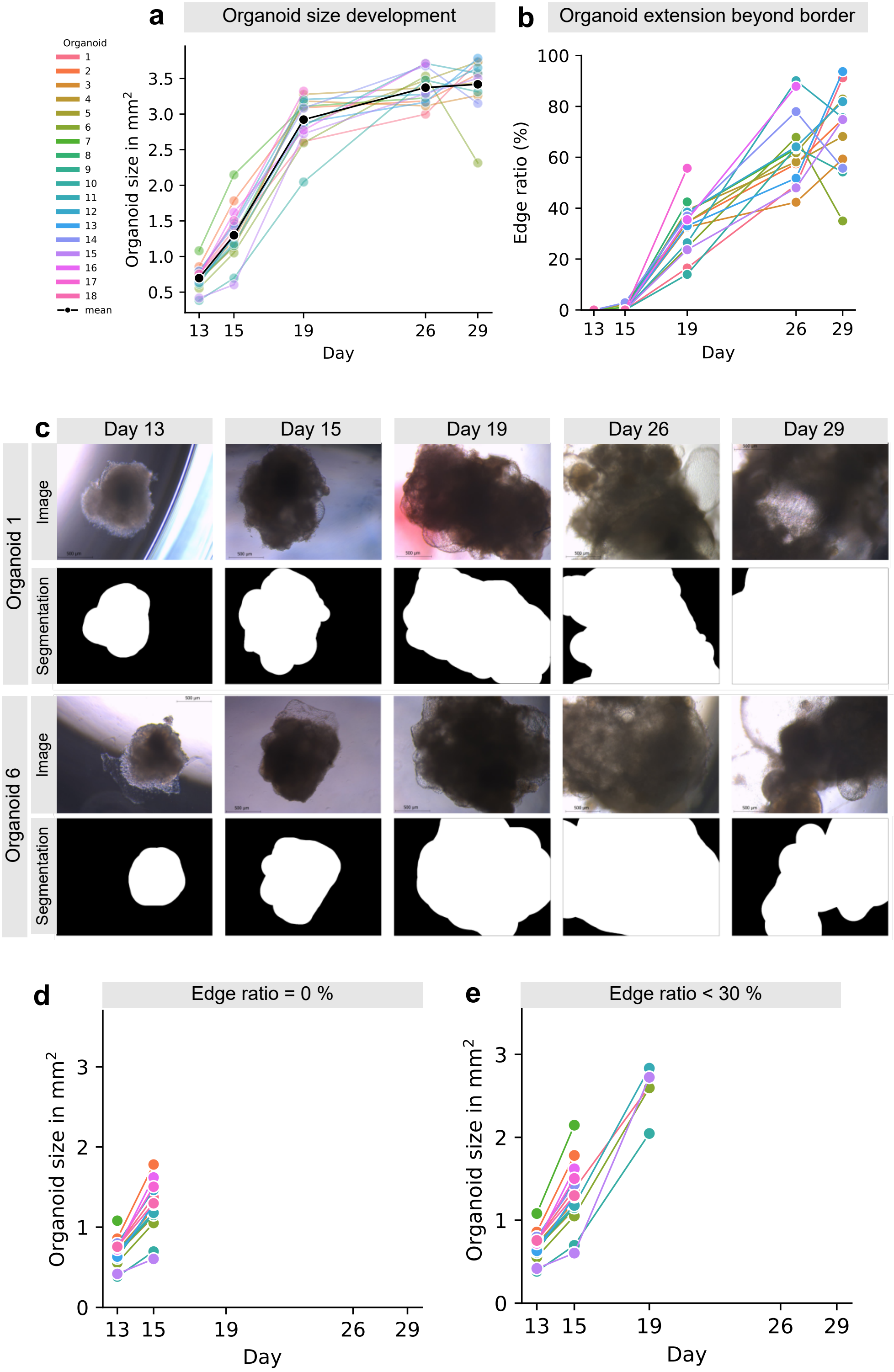


Figure 2.19: Concept of automated detection and handling of organoid extensions beyond imaging boundaries. (a) Development of organoid size for 18 wildtype brain organoids from day 13 to day 29. (b) Percentage of segmentation pixels at the image edge classified as organoid (*Edge ratio*), as defined in Equation 2.11. (c) Images and corresponding segmentations of organoids 1 (top) and 6 (bottom) across days 13 to day 29. (d-e) Organoid size for cases with an Edge ratio 0% (d) and <30% (e).

the imaging field compared to a lower ratio. This could also occur if the organoid is not centered within the imaging field and large organoid protrusions extend far beyond the imaging border (Figure 2.20a).

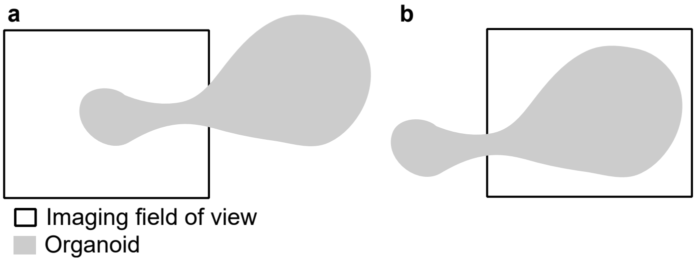


Figure 2.20: Organoid centering in imaging field of view affects the interpretability of the Edge ratio. (a) If an organoid is off-center and exhibits large protrusions extending beyond the imaging border, a low Edge ratio does not necessarily imply that most of the organoid is contained within the field of view. (b) A well-centered organoid yields a more reliable and interpretable Edge ratio.

To ensure complete imaging, it is recommended to position organoids at the center of the imaging field and to select an appropriate magnification that accommodates the entire organoid. These adjustments help avoid imaging artifacts and ensure accurate segmentation. The Edge ratio serves as an effective tool for automatically detecting organoid extensions beyond imaging boundaries, enabling the flexible exclusion of organoids based on a defined threshold of boundary contact.

2.2.5. Discussion

Brightfield microscopy continues to be a pivotal method for monitoring the growth and size of brain organoids [61, 150, 198, 201], particularly given its accessibility and the feasibility compared to more complex techniques like MRI (Section 2.1). However, while this method remains widely adopted, several limitations persist, particularly in terms of manual analysis, which can introduce substantial observer bias and time inefficiency when dealing with large sample sizes [202]. This study addressed these limitations by evaluating and comparing a variety of image analysis pipelines for organoid segmentation, growth monitoring, and clone diversity assessment, focusing on both classical and deep learning-based methods.

The deep learning-based SegFormer [111], applied here for the first time in the context of organoid segmentation, demonstrated superior segmentation accuracy, excelling in robustness and generalizability across labs and conditions. It consistently outperformed classical methods like CellProfiler [130] and OrganoSeg [150], as well as the deep learning-based MOrgAna [166], specifically designed for brain organoids, achieving higher Dice scores across developmental stages. While SegFormer adapted well to varying experimental conditions and imaging artifacts, MOrgAna showed greater variability, particularly in Lab B, emphasizing the need for models that can reliably generalize across datasets.

The likely reason for MOrgAna's weaker performance lies in its architectural design. It utilizes a multilayer perceptron (MLP) with two hidden layers for segmentation, classifying it as a deep learning model. However, unlike SegFormer, which is a transformer-based model optimized for capturing spatial hierarchies, MOrgAna relies solely on fully connected layers to process pixel-wise information. While MLPs excel at learning complex non-linear mappings, they cannot inherently capture spatial dependencies, a key advantage of convolutional and transformer-based neural networks [203].

In terms of organoid diversity, the analysis demonstrated the utility of feature extraction methods to quantify clone-specific morphological characteristics. The use of PCA to visualize clone diversity revealed distinct morphological differences among the clones, further supporting the value of automated analysis in identifying subtle phenotypic variations that may not be easily observable through manual inspection.

Optical distortions from well transfers and Matrigel embedding, along with organoids extending beyond imaging borders, can confound size measurements and growth analyses. Standard protocols do not provide intrinsic quality control to detect or correct such issues. Distortions from different well types and Matrigel embedding can cause apparent enlargement, complicating accurate quantification. Comparing pre- and post-transfer measurements allowed these effects to be identified and corrected, improving growth tracking. Additionally, as organoids expand beyond imaging boundaries, accurate monitoring becomes crucial. The Edge ratio metric enables the detection and exclusion of incomplete data, ensuring only fully imaged organoids contribute to analyses, reducing bias and enhancing reliability.

In conclusion, this work emphasizes the importance of selecting appropriate image analysis tools and protocols for brain organoid research. The findings demonstrate that deep learning-based methods, particularly SegFormer, offer significant advantages over classical techniques in terms of segmentation accuracy, robustness, and ability to handle diverse experimental conditions. Furthermore, careful attention to factors such as optical distortion and imaging boundaries can greatly enhance the reliability of organoid growth monitoring. Future studies should continue to refine these methods, exploring their application across different types of stem cell models and experimental conditions.

2.3. Epifluorescence Microscopy

> This section is based on brain organoid images cultivated and imaged by Sabine Jung-Klawitter, Julian Schröter, and Petra Richter. S.J.-K. contributed to nuclei annotation, while she, J.S., and L.D. performed ventricle annotation. L.D. was responsible for data processing, analysis, and the development of deep learning models.

Compared to brightfield microscopy, epifluorescence microscopy enables more detailed structural analysis of brain organoids while remaining more accessible than light-sheet imaging.

This section outlines a conceptual framework for deep learning-based analysis of immunohistochemistry-stained epifluorescence images, aiming to segment nuclei and ventricle-like structures as a foundation for assessing neuronal development (Figure 2.21, Section 2.3.1, Section 2.3.2). High-quality images suitable for annotating neurons were not available at the time of thesis submission, and thus, this component is deferred to future studies. In addition, Section 2.3.3 explores registration-based 3D reconstruction from consecutive slices, demonstrating its feasibility while identifying key challenges for automated volumetric analysis.

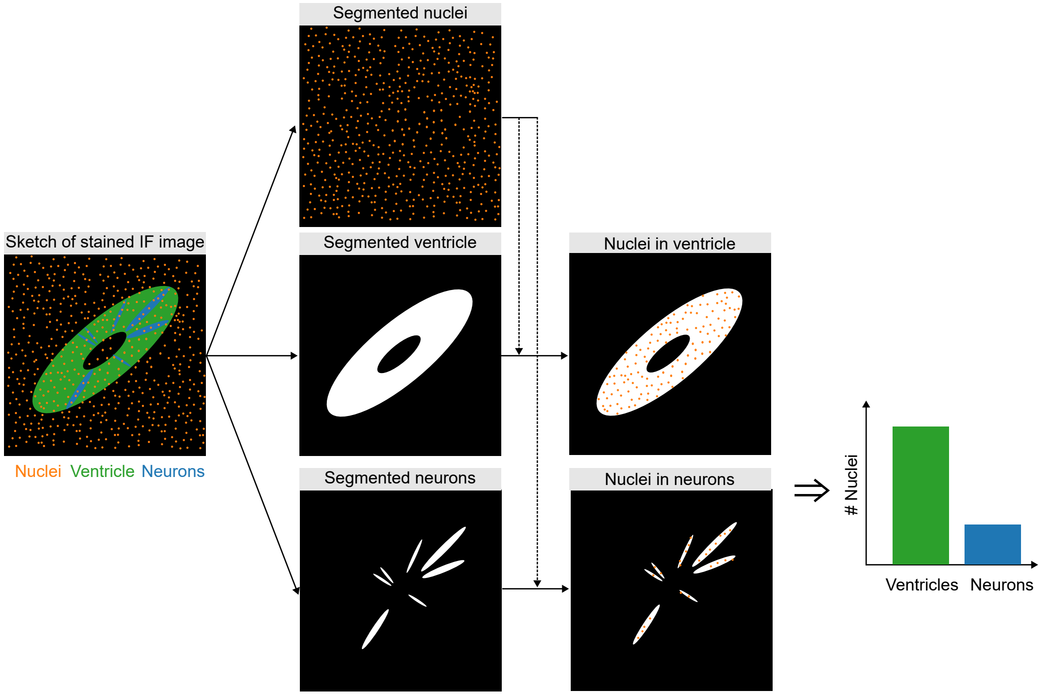


Figure 2.21: Concept for measuring brain organoid neuronal development progress using epifluorescence microscopy. The idea is to segment nuclei, ventricles, and neurons. The ratio of nuclei within ventricles to those within neurons serves as an estimate for the progression of neuronal development.

2.3.1. Nuclei Segmentation

2.3.1.1. Data

For nuclei segmentation, the data used in this study consisted of a DAPI-stained epifluorescence image of a cerebral organoid slice (20–30 μm thick, 900 × 900 px), acquired through immunohistochemistry and imaged with an epifluorescence microscope (Figure 2.22a). DAPI staining specifically labels the nuclei. The image was manually annotated to mark the positions of nuclei by setting a single pixel within each visible nucleus (Figure 2.22b). The annotated image was then divided into non-overlapping 32 × 32 px patches, with only patches containing at least one annotated nucleus included for model training.

Two datasets were created for model evaluation. Dataset A includes all patches containing at least one annotated nucleus, regardless of nuclei boundary clarity (Figure 2.22c,d). Dataset B is a subset of Dataset A, containing empirically selected patches with relatively clear nuclear boundaries (Figure 2.22c,e). Dataset B was created to serve as a reference for well-segmentable cases, allowing for a distinction between straightforward and challenging segmentation scenarios.

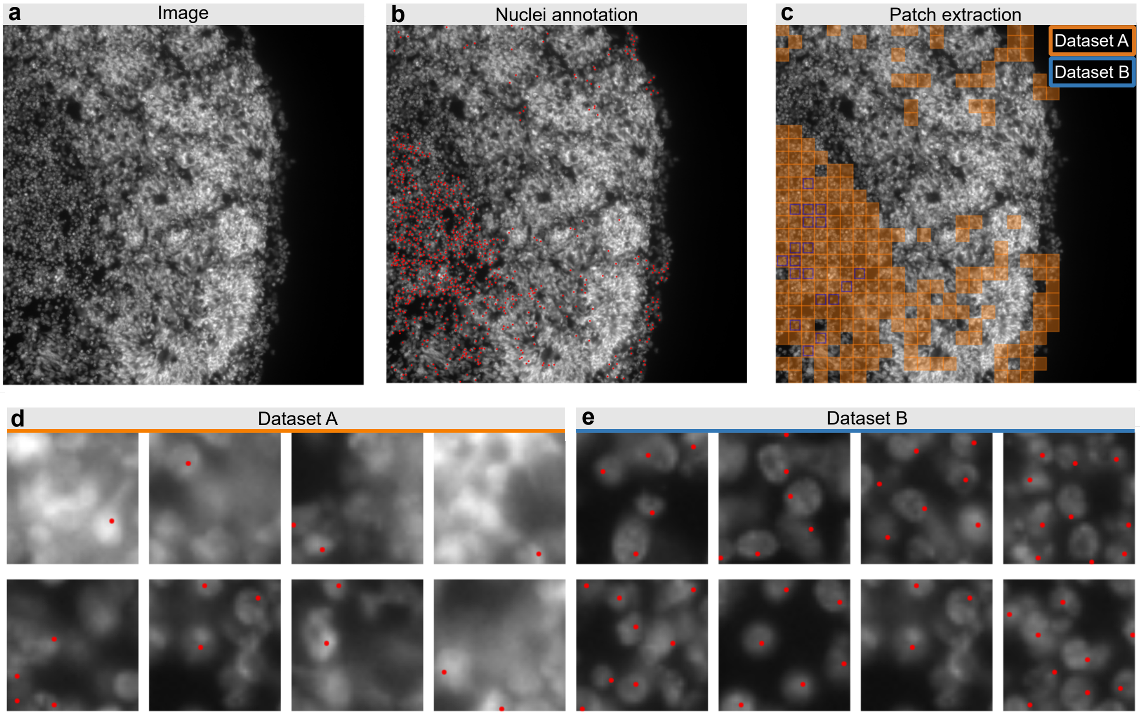


Figure 2.22: Data preparation and extracted patches for nuclei segmentation. (a) DAPI-stained slice of a cerebral organoid. (b) The same image from (a), overlaid with nuclei annotations. (c) The image from (a) with the locations of extracted patches indicated: orange marks for Dataset A and blue marks for Dataset B, where Dataset B is a subset of Dataset A. (d-e) Example patches extracted for Dataset A and Dataset B.

2.3.1.2. Methods

For automated nuclei segmentation, BeadNet [108], a deep learning-based segmentation model, was trained with its default configuration separately on manually annotated Dataset A and Dataset B. Although pre-trained BeadNet models are available, retraining was necessary due to differences in image characteristics. The training setup included a batch size of 2, a learning rate of 0.006, and convolutional layers with filter sizes of 64, 128, 256, and 512. Training was performed for up to 150 epochs, with early stopping applied if no improvement was observed over 30 consecutive epochs. Data augmentation included BeadNet-specific implementations of image flipping, scaling, rotation, contrast adjustment, blurring, and noise addition to enhance robustness and generalization.

To assess the variability in model performance, five BeadNet models were trained, each with a unique random initialization. The datasets were split into 60% for training, 20% for validation, and 20% for testing.

Model evaluation was performed using the F1-score-inspired metric Q_F (Equation 2.14), introduced by [108], which quantifies segmentation performance in the BeadNet framework. A perfect segmentation results in a Q_F of 1, while the worst possible performance yields a score of 0. Even though nuclei are annotated as single-pixel points, exact localization is not required for a prediction to be considered correct. Instead, a prediction is considered correct if it lies within a defined radius around the ground truth point. To accommodate this tolerance, the ground truth annotations are dilated in two steps: first with a cross-shaped, then with a square-shaped structuring element, as described in [108]. In the following, n denotes the number of test images i in the dataset:

$$\begin{aligned}
Q_\mathrm{F} &= \frac{1}{n}\sum_{i=1}^{n}\frac{2Q_{\mathrm{P},i}Q_{\mathrm{R},i}}{Q_{\mathrm{P},i}+Q_{\mathrm{R},i}} \\
Q_{\mathrm{P},i} &= \frac{1}{n}\sum_{i=1}^{n}\frac{tp_i}{tp_i+fp_i} \\
Q_{\mathrm{R},i} &= \frac{1}{n}\sum_{i=1}^{n}\frac{tp_i}{tp_i+fn_i} \\
tp_i &:= N_{\mathrm{pred},i}-N_{\mathrm{split},i}-N_{\mathrm{add},i} \\
fp_i &:= N_{\mathrm{split},i}+N_{\mathrm{add},i} \\
fn_i &:= N_{\mathrm{miss},i}
\end{aligned} \tag{2.14}$$

$N_{\mathrm{split},i}$ is the number of *split* nuclei (multiple predicted nuclei lie in a ground truth nucleus) in test image i. $N_{\mathrm{miss},i}$ is the number of missing nuclei (no predicted nuclei lies in a ground truth nuclei), and $N_{\mathrm{add},i}$ is the number of added nuclei (a predicted nuclei lies in ground truth background).

2.3.1.3. Results

Inspection of the annotations revealed a concentration of annotated nuclei in the lower-left region of the image, an area characterized by lower illumination but with nuclei that appear better separated (Figure 2.22b). In contrast, the right side of the image contains few annotated nuclei, despite higher illumination. In this region, nuclei are less clearly distinguishable and generally more sparsely distributed compared to the left. This observation highlights the need to refine imaging and annotation protocols, as even manual segmentations become inconsistent in more ambiguous regions.

BeadNet was trained and tested on datasets A and B in a cross-testing setup to evaluate the model's capabilities. Dataset A consisted of all patches, whereas Dataset B was a curated subset of A, comprising patches with more clearly visible nuclear boundaries. When trained and tested on Dataset A, the model performed poorly (Table 2.6). This indicates that the model struggled to reproduce the annotation results on Dataset A. Conversely, the model trained and tested on Dataset B demonstrated strong performance on this curated subset (Table 2.6). When the same model trained on Dataset B was tested on Dataset A, Q_F dropped (Table 2.6). These substantial fluctuations in performance suggest that the model is capable of achieving reliable results on data with clearly defined nuclear boundaries (Figure 2.22e).

Table 2.6: Cross-testing. Q_F is presented as mean ± SD.

Exp	$\text{Dataset}_{\text{Train}}$	$\text{Dataset}_{\text{Test}}$	Q_F
#1	A	A	0.09 ± 0.04
#2	A	B	0.03 ± 0.05
#3	B	A	0.13 ± 0.02
#4	B	B	0.85 ± 0.03

Qualitative evaluation shows that the model trained on Dataset A predicts only a few nuclei in both regions with relatively clear and blurry nuclear boundaries (Figure 2.23a,b). This may be attributed to the limited number of annotated nuclei in blurry regions, leading the model to adopt a generally conservative prediction strategy. In contrast, the model trained on Dataset B produces substantially more predictions in both clearly and poorly defined regions (Figure 2.23c). Exposure to well-defined nuclei during training likely increases the model's confidence. Compared to the ground truth, predictions from the Dataset B model appear more centered within nuclei and more uniformly distributed across the field of view (Figure 2.23a,c).

It is hypothesized that the discrepancy in model performances is due to the subjective and noisy annotations present in Dataset A. The underlying cause is likely the use of overly thick organoid slices, which result in nuclei overlapping in the z-direction and blurred nuclear boundaries (Figure 2.22d). To improve data quality and model performance, it is recommended that thinner organoid slices be acquired or z-stacks be recorded to resolve nuclei boundaries more clearly.

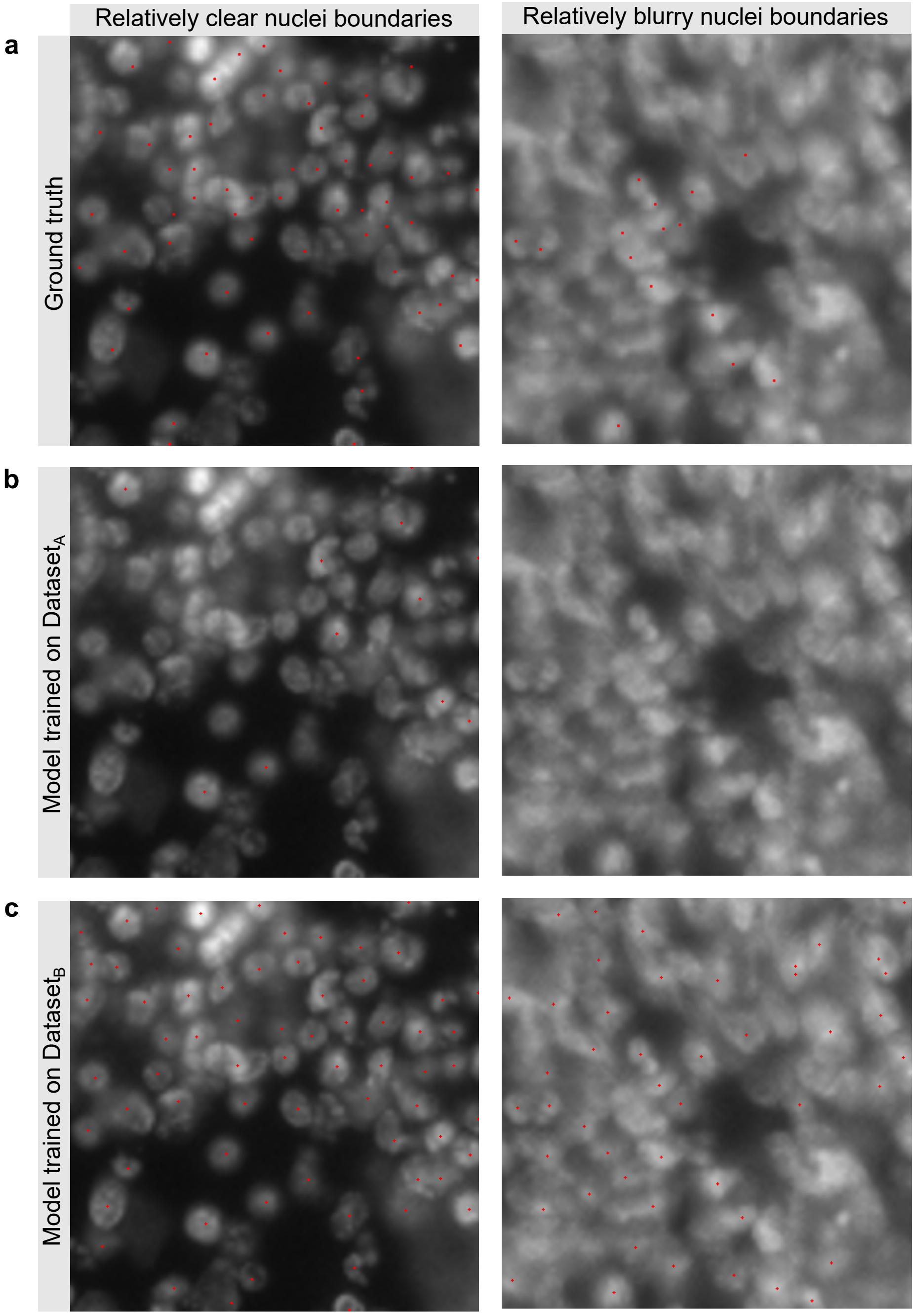


Figure 2.23: Model predictions from training on Dataset A and Dataset B across regions with varying degrees of nuclear blurring. (a) Ground truth nucleus annotations. (b) Predictions from model trained on Dataset A. (c) Predictions from model trained on Dataset B. The left column represents a region with well-defined nuclear boundaries, while the right column shows a region with blurry boundaries, allowing comparison of model performance under different conditions. Both image patches are taken from test regions not seen during training.

A promising approach to estimating nuclei in regions with blurry nuclei boundaries involves training BeadNet on patches with well-defined nuclear boundaries and subsequently applying the trained model to new images, including regions with blurry nuclear boundaries. The observations indicate that raw BeadNet predictions encode valuable information about potential nuclei positions (Figure 2.24). By manually adapting the threshold parameters for each image based on the model's predictions, the approach enabled the identification of candidate nuclei even in areas with poorly defined boundaries. As an alternative to manual adjustment, the threshold could be estimated by specifying an approximate number of expected nuclei per image and iteratively lowering the prediction threshold until that number is reached. The primary challenge lies in quantitative evaluation, as ground truth annotations are unavailable for such regions. Nonetheless, this approach holds the potential for a semi-supervised tool for nuclei counting.

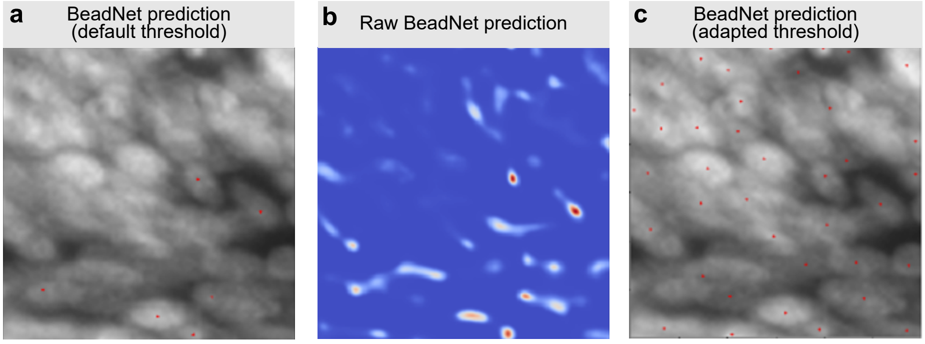


Figure 2.24: Raw BeadNet prediction encodes information about potential nuclei positions. (a) DAPI-stained cerebral organoid slice overlaid with BeadNet predictions using the default threshold. (b) Raw BeadNet predictions corresponding to the overlay in (a). (c) BeadNet predictions after applying a manually adjusted threshold, highlighting additional potential nuclei positions.

2.3.1.4. Protocol Optimization

BeadNet demonstrated reliable performance in cases with clear nuclear boundaries. Since clear nuclear boundaries can be influenced by organoid slice thickness, future data generation efforts could focus on capturing z-stacks with a thickness approximating a single cell diameter. This approach would help reduce nuclei stacking along the z-axis, improving nuclei visibility and boundary delineation.

Automated quantification of protocol improvements could utilize BeadNet with ensemble-based uncertainty to identify regions of inconsistent segmentation, highlighting data or protocol limitations. This approach could guide iterative optimizations, such as adjusting slice thickness or staining protocols.

2.3.2. Ventricle Segmentation

2.3.2.1. Data

The data used in this study were acquired from immunohistochemistry preparations of cerebral organoid slices (20–30 μm thick) imaged using an epifluorescence microscope. This included images from 39 cerebral organoid slices, each stained with SOX2, a marker for ventricle-like structures. Each slice originated from a distinct organoid. Since the data was collected from multiple experiments and imaged by different individuals, the images varied: some depicted the entire cerebral organoid, while others focused on specific areas displaying ventricle-like structures within the organoid.

To contextualize model performance relative to human experts, annotations were provided by three annotators (A1, A2, A3). Additionally, a consensus annotation was generated using a pixel-wise majority vote among the three annotators.

2.3.2.2. Methods

For ventricle segmentation, binary semantic segmentation was performed to separate ventricles from the microscopy background or other organoid tissue. For automated semantic segmentation, two methods were used: Otsu's thresholding [105] and a SegFormer model [111]. The SegFormer model with the MiT-B2 encoder configuration was used. All processes for model training, evaluation, and inference utilized the implementation from [184]. The model was trained with the AdamW optimizer, configured with a learning rate of 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 0.1, using a batch size of 2 for a total of 5,000 iterations. A weighted combination of binary cross-entropy and Dice loss (ratio 1:10) was employed to guide the learning process (Equation 1.8). On-the-fly image augmentation included four steps:

1. Image downscaling to 512 × 512 px.
2. Random horizontal flipping with a probability of 0.5.
3. Z-score normalization (Equation 2.10).

The model was evaluated using the Dice score (Equation 1.6) with leave-one-out cross-validation (Section 1.4.5). Separate models were trained for each annotator, where M_{A1}-M_{A3} denote models trained on annotations from A1–A3, and M_{AC} refers to the model trained on the consensus annotation derived from the three annotators.

2.3.2.3. Results

The results show that the SegFormer performed substantially better compared to Otsu's thresholding (Table 2.7, Figure 2.25a). The highest performance was consistently achieved when the SegFormer model was trained and evaluated on the same annotation. For instance, M_{A1} performed best when evaluated on A1 (Table 2.7, Figure 2.25a).

This result is expected, as models trained and tested on the same annotation also learn annotator-specific features. However, a key question is how well these models generalize across different annotations. For example, how does M_{A1} perform on A2 and A3? The results showed that SegFormer models tested on annotations different from those they were trained on achieved similar performance to the respective human annotators. For instance, there was no significant performance difference between A1 and M_{A1} when evaluated on A2 and A3 (Figure 2.25a). This suggests that the models performed on par with the human annotators they were trained on.

Achieving human-level performance is already beneficial, as it allows for more efficient analysis without sacrificing quality. Yet, an even more critical question is whether a model can surpass human annotators, for example, by offering greater consistency. Across all annotation sets, the model trained on the consensus annotation (M_{AC}) consistently ranked second-best (Table 2.7) and slightly outperformed individual annotators (Figure 2.25b). This indicates that the consensus-trained model exceeds individual human annotators.

Table 2.7: Model and annotator performance for ventricle segmentation across multiple annotators. The first three rows show pairwise agreement between human annotators (A1–A3). The remaining rows report the performance of automated methods evaluated against each annotator. M_{A1}-M_{A3}: SegFormer trained on annotations A1-A3. M_{AC}: SegFormer trained on the consensus of the three annotators. Otsu, Otsu's thresholding. All values are reported as Dice scores (mean±SD).

Annotator / Method	**A1**	**A2**	**A3**
A1	-	0.79 ± 0.19	0.77 ± 0.14
A2	0.79 ± 0.19	-	0.76 ± 0.18
A3	0.77 ± 0.14	0.76 ± 0.18	-
Otsu	0.38 ± 0.20	0.37 ± 0.21	0.37 ± 0.20
M_{A1}	0.90 ± 0.10	0.76 ± 0.19	0.76 ± 0.14
M_{A2}	0.75 ± 0.21	0.90 ± 0.15	0.72 ± 0.20
M_{A3}	0.73 ± 0.17	0.72 ± 0.19	0.88 ± 0.16
M_{AC}	0.83 ± 0.13	0.82 ± 0.17	0.80 ± 0.15

2.3.2.4. Protocol Optimization

Imaging

Since the data was obtained from various experiments and imaged by different individuals, the images varied in scope—some captured the entire cerebral organoid, while others focused on specific regions, particularly ventricle-like structures. For more conclusive comparative analyses, acquiring full-slide microscopy images is essential for two key reasons: (i) Imaging only parts of the organoid introduces observer bias, as selected areas may not accurately represent the entire organoid. This bias cannot be corrected *post hoc* due to the absence of full organoid images. (ii) Partial visibility of ventricles in some images complicates the objective assessment of their presence. Full-slide imaging guarantees

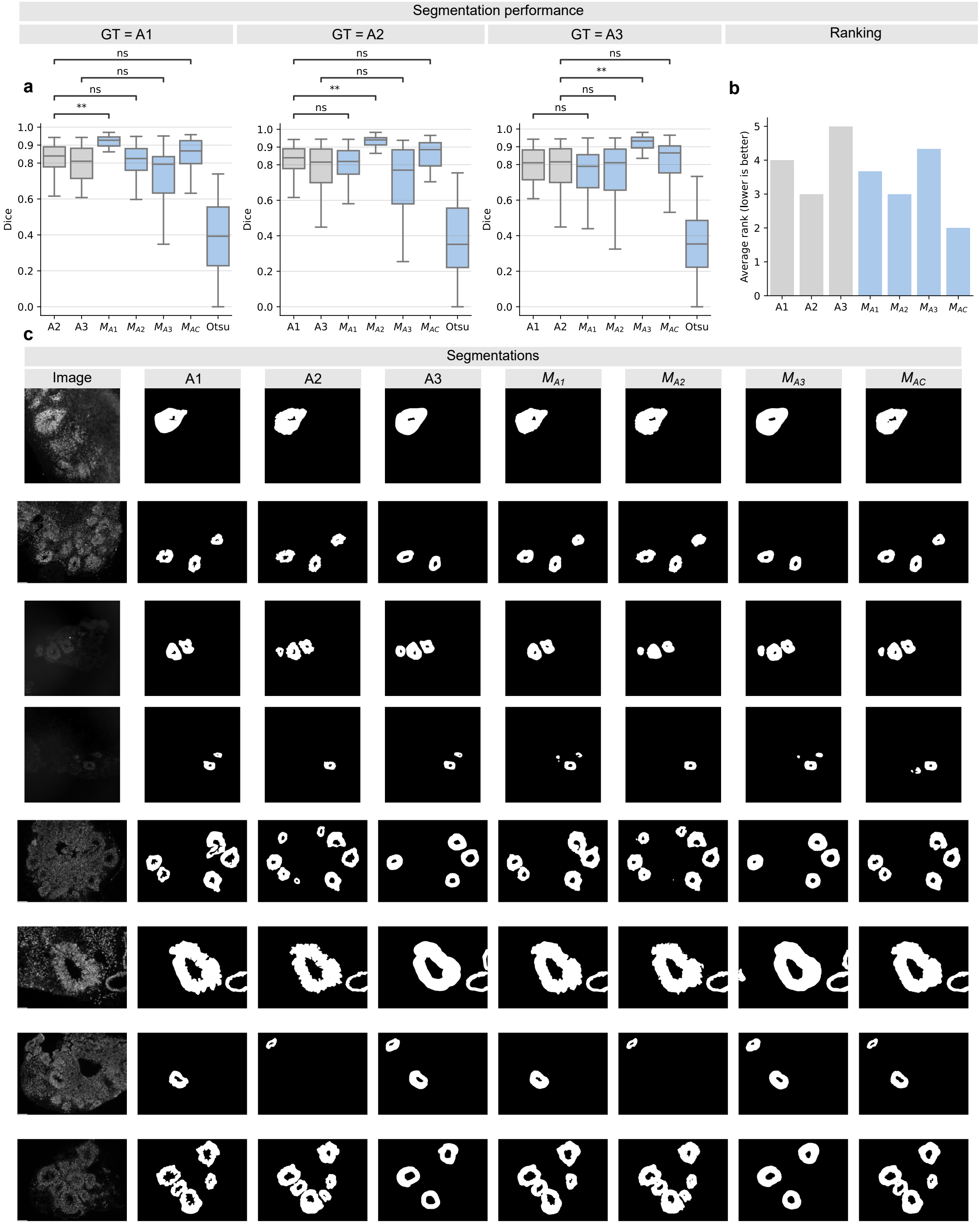


Figure 2.25: Ventricle segmentation performance across multiple annotators. M_{A1}-M_{A3}: SegFormer trained on A1-A3. M_{AC}: SegFormer trained on the consensus of three annotators. Otsu: Otsu's thresholding. (b) is derived from the results in (a), ranking all models based on their mean Dice score.

more consistent and comprehensive data, reducing bias and improving the accuracy of organoid analyses.

Staining

To further enhance segmentation performance, improving the specificity of ventricle-like structure stainings is essential. Unspecific staining, which increases the visible tissue area, is hypothesized to reduce annotation certainty. Annotation uncertainty among experts is likely indicative of ambiguity in organoid analyses, which should be minimized whenever possible. To evaluate this hypothesis, annotator and model certainty were calculated for all images i of the image dataset I. Similar to prior work on model uncertainty estimation [87, 204], a certainty vector $\boldsymbol{C} \in \mathbb{R}^{|I|}$ was defined, where each component C_i represents the certainty for a specific image $i \in I$, calculated as the mean Dice score across all pairs of mask sets $(j,k) \in P$:

$$\boldsymbol{C} = [C_i]_{i \in I}, \quad C_i = \frac{1}{|P|} \sum_{(j,k) \in P} \mathrm{Dice}(M_{i,j}, M_{i,k}) \tag{2.15}$$

where:

- $M_{i,j}$: Segmentation mask corresponding to image i and annotation or model j
- P: Set of all unique segmentation pairs from a available segmentations
- $|P| = \binom{a}{2} = \frac{a(a-1)}{2}$: Number of unique segmentation pairs

The definition of P depends on the certainty type:

$$\text{Annotator Certainty:} \quad P = \{(A_1, A_2), (A_1, A_3), (A_2, A_3)\} \tag{2.16}$$

$$\text{Model Certainty:} \quad P = \{(M_{A1}, M_{A2}), (M_{A1}, M_{A3}), (M_{A2}, M_{A3})\} \tag{2.17}$$

Higher values of C_i indicate greater agreement among segmentations, and thus higher certainty. This metric enables quantification of ambiguity in annotations or model predictions at the image level.

The analysis revealed a Pearson correlation of -0.36 ($p = 0.02$) between annotator certainty and tissue area (determined using Otsu's threshold). This moderate negative correlation indicates that as tissue area in microscopy images increases, annotation certainty decreases. Enhancing marker specificity could thus reduce annotator variability. Additionally, model certainty showed a strong positive correlation with annotation certainty ($r = 0.83, p = 5.6 \times 10^{-11}$).

These findings indicate that (i) Unspecific stainings and greater tissue visibility contribute to higher annotator variability, and (ii) Increased annotator variability leads to higher variability in automated segmentation. Improving stain specificity and imaging entire organoids are therefore crucial steps toward reducing variability and enhancing both manual and automated segmentation quality.

Model certainty can be leveraged for protocol optimization, such as selecting markers for organoid ventricle staining, provided the model was previously trained to detect the same substructure (e.g., ventricle-like regions). In such cases, the model could be applied in a one-shot manner to guide marker selection. In principle, this approach could be extended to other aspects of protocol design, including organoid slicing. For organoid staining, applying all markers to the same organoids is preferred (Figure 2.26a), because it avoids issues with organoid batch-to-batch variability when comparing model certainty across different biological samples. For each marker, predicted segmentation masks can be used to calculate model certainty (Figure 2.26b, Equation 2.15, 2.17). By comparing model certainties across markers, the marker with the highest certainty can be identified and selected for subsequent comparative analyses of organoids (Figure 2.26c).

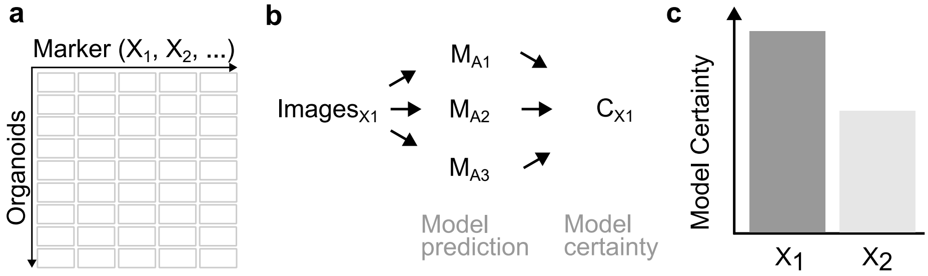


Figure 2.26: Concept of staining-marker optimization using model certainty. (a) Organoid stainings with different markers. (b) Calculation of marker-specific model certainty using Equations 2.15 and 2.17. (c) The marker showing the highest model certainty is the optimal marker.

2.3.3. Registration-Based Volumetric Reconstruction

Previous measurements were based on 2D organoid slices. However, volumetric measurements of organoids would be beneficial, providing a more complete and comprehensive perspective on the global cerebral organoid organization. This section evaluates the feasibility and challenges of *in silico* volumetric reconstruction based on serial epifluorescence microscopy slices of cerebral organoids.

2.3.3.1. Data and Preprocessing

DAPI-stained epifluorescence microscopy images from 54 serial slices of the same cerebral organoid were utilized. Registration-based volumetric reconstruction of cerebral organoids presents multiple challenges. To assess feasibility, key issues were first addressed manually. Sections 2.3.3.4 and 2.3.3.5 examine the potential for automated detection and correction of these challenges, and discuss how they can inform the development of criteria for iterative protocol optimization.

One challenge is tissue stretching, which requires affine transformations to achieve accurate alignment (Figure 2.27a). Another complication arises from significant tissue rotations (Figure 2.27b), which can misalign features and cause registration algorithms to

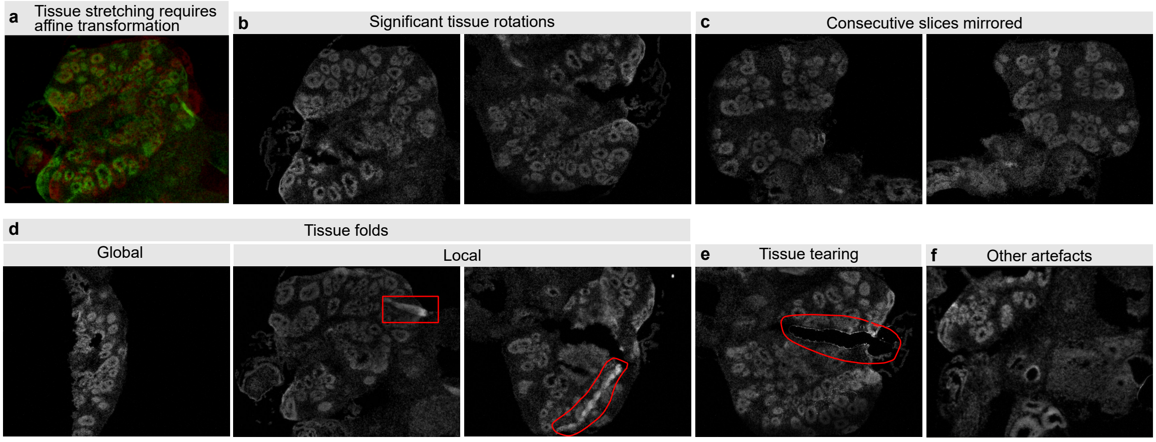


Figure 2.27: Challenges in registration-based volumetric reconstruction of cerebral organoids. (a) Tissue stretching during sample preparation necessitates affine transformations for accurate alignment. (b,c) Tissue may be significantly rotated or mirrored between consecutive sections due to sample handling. (d) Global and local tissue folding can obstruct the visibility of the whole tissue or specific regions. (e) Tissue tearing hampers registration, as corresponding substructures no longer align. (f) Other artifacts include out-of-focus sections, combinations of the aforementioned challenges, or slices that do not match adjacent slices.

converge on incorrect solutions. This issue can be mitigated by applying pre-processing steps, such as manual or automated rotation correction, to roughly align the images beforehand. In this study, larger rotations were manually identified for each image pair, and the corresponding moving image (i.e., the image to be registered) was rotated accordingly before registration.

Sample mirroring (Figure 2.27c) introduces similar challenges. Mirrored slices can mislead the registration algorithm, again resulting in poor alignment. Observations suggest that mirroring often affects multiple consecutive slices, likely due to sample handling during imaging. This results in distinct mirroring events, after which all subsequent slices remain mirrored until another such event occurs. For manual correction of sample mirroring, all images should be screened at the start of the analysis to detect the onset of mirroring. From that point onward, slices can be systematically corrected using *in silico* mirroring to ensure proper alignment. Additionally, when mirroring occurs, the side of the sample with a greater number of available slices should be prioritized for further processing. This helps minimize the impact on data integrity, as *in silico* mirroring may not perfectly replicate the true appearance of the sample. If consecutive slices were observed to be mirrored, the moving image was manually mirrored before registration.

Tissue folding is another common issue, affecting either large regions (global folds) or smaller areas (local folds) of the tissue (Figure 2.27d). These folds can compromise data quality by obscuring important regions and introducing artifacts. Global folds, which obscure larger portions of the tissue, should generally be excluded from the analysis to maintain data quality. Local tissue folds, which affect smaller regions, may be retained or excluded depending on their severity and impact.

Additionally, tissue tearing (Figure 2.27e) and out-of-focus sections or unexplained artifacts that do not correspond to adjacent slices (Figure 2.27f) are challenges for registration. Depending on the severity of the tear and blurriness, affected slices may need to be excluded. Before image registration, one global tissue fold, one out-of-focus image, and one artifact caused by tissue tearing were manually excluded.

2.3.3.2. Methods

The volumetric reconstruction of the cerebral organoid was performed through iterative pairwise image registration using the Python ITK elastix library [113] (version 0.18.1). The procedure involved registering consecutive 2D image slices from DAPI-stained microscopy images and aligning each slice to the previous one to correct for translational, rotational, and affine displacements. The process began by selecting the first image slice as the fixed reference. Subsequent slices were then sequentially aligned to the fixed reference through pairwise registration using the ITK function *elastix_registration_method*, with each moving image being registered to the previously aligned (fixed) image.

The registration process utilized a multi-resolution approach with ten levels of resolution and incorporated three sequential transformation steps. These steps included: translation, to correct for simple shifts between images; rigid transformation, to align images by addressing rotation and scaling; and affine transformation, to refine alignment by accounting for tissue skew and deformation (Figure 2.27a). For all three transformation steps, the default parameters provided by ITK elastix were applied. After each registration step, the newly registered image became the fixed reference for the next iteration. Upon completing the iterative process, all registered images were stacked to create a 3D volumetric reconstruction of the cerebral organoid.

2.3.3.3. Results

Registered images demonstrated improved alignment of ventricle-like structures compared to non-registered images (Figure 2.28). However, certain challenges remained, such as local tissue folds (Figure 2.28, images 4 and 6) and tissue skew (Figure 2.28, image 5, bottom-left), which resulted in substructures that did not perfectly overlap.

2.3.3.4. Pipeline Automation

Due to the availability of only one biological sample, manual pre-registration and slice mirroring were employed to obtain initial meaningful results. However, to transition toward a fully automated analysis pipeline, further considerations are necessary to address challenges that were previously resolved manually.

Significant tissue rotations can pose challenges for registration, as they can misalign features and lead to convergence on incorrect solutions. An idea for automating the detection and correction of rotation-induced registration issues involves identifying the

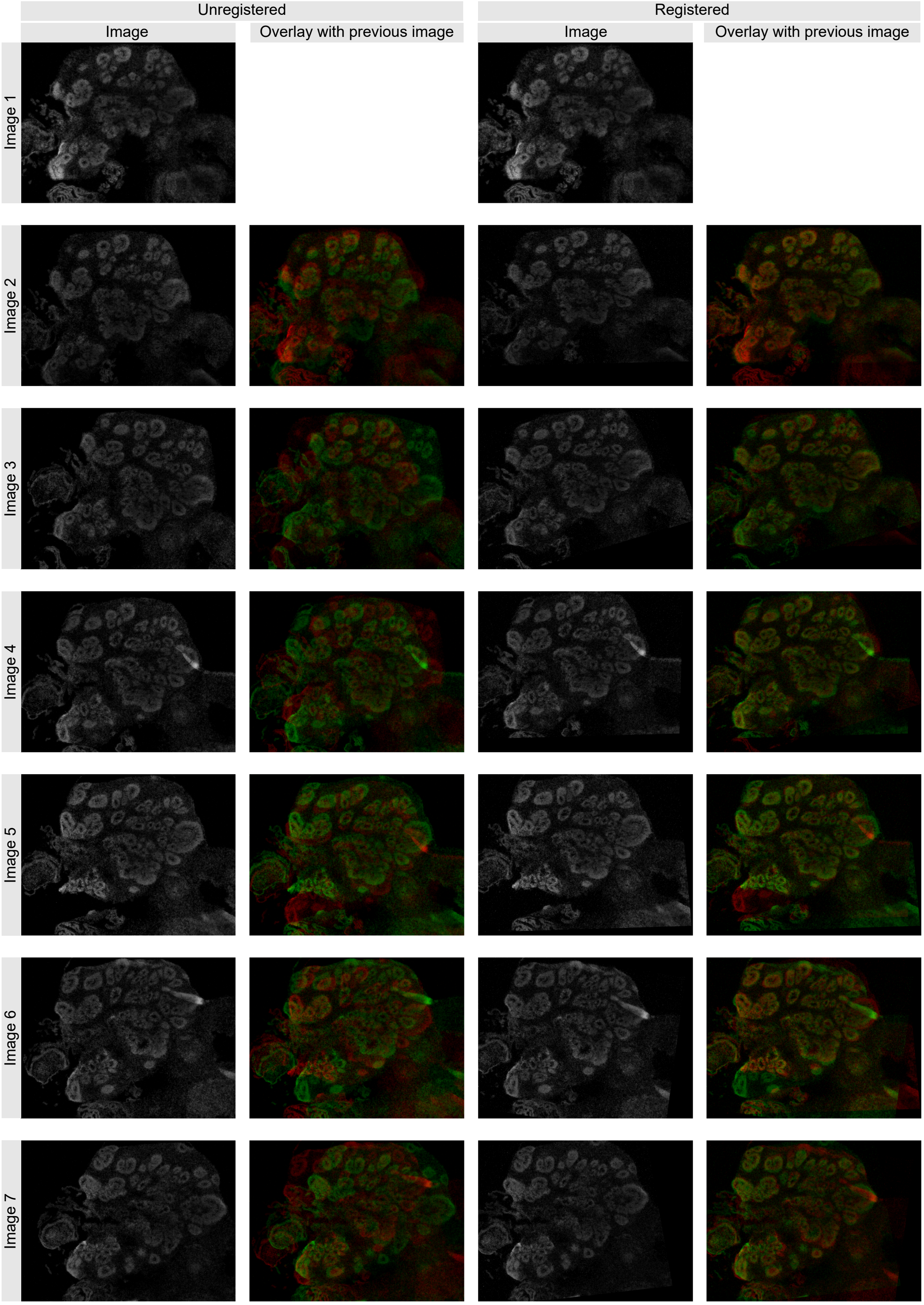


Figure 2.28: Consecutive unregistered and registered tissue sections. Consecutive sections are shown from top to bottom. The first column displays the original unregistered images, while the third column presents the registered images. The second and fourth columns provide overlays of the fixed image (previous image, red) and the moving image (current image, green) to visually assess the registration quality.

optimal pre-registration angle by minimizing the registration error across multiple tested rotation angles (Figure 2.29). An analogous strategy can be applied to handle sample mirroring in consecutive slices.

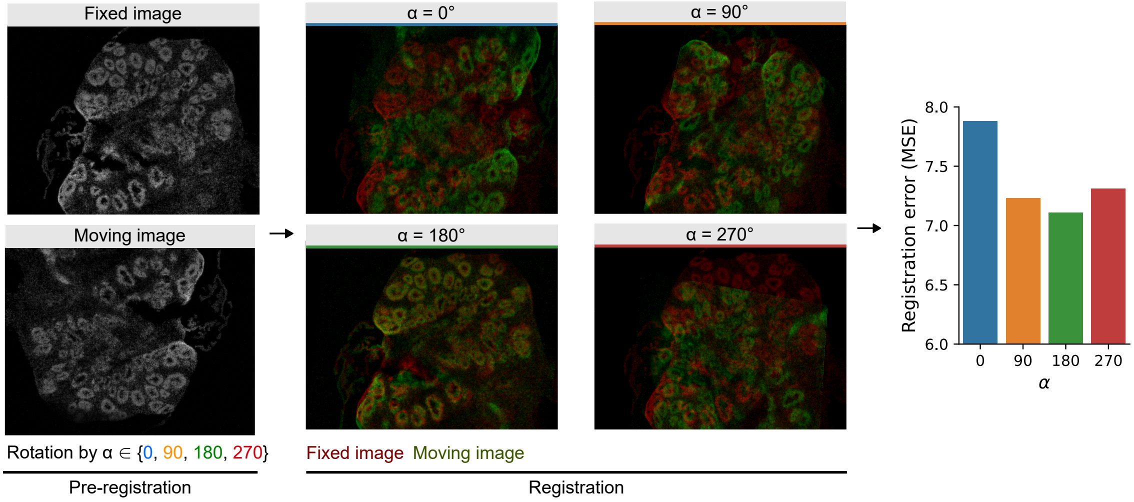


Figure 2.29: Example for registration error-guided correction of rotation-induced registration issue.

This registration error-based approach can also be extended to detect global tissue folds and tissue tears. Elevated registration errors can indicate the presence of such folds, which may require exclusion from further analysis. Alternatively, deep learning provides promising tools for automating the detection and handling of tissue artifacts. For instance, models could be trained for binary classification to identify and exclude slices affected by global folds or tearing, or for semantic segmentation to precisely delineate regions with local folds or tissue tearing.

Currently, the detection and correction of these artifacts are performed manually. Moving toward full automation will require the generation of a larger, annotated dataset to train and validate machine learning models, as well as a more comprehensive understanding of the variability and diversity of artifacts present in larger datasets.

2.3.3.5. Protocol Optimization

Provided that additional annotations and segmentations are available, the suggested methods could be trained to automatically evaluate overall data quality by quantifying the following criteria:

- Frequency of global tissue folds
- Frequency, absolute area, and relative area of local tissue folds
- Frequency of out-of-focus images
- Frequency of mirroring events

These quantitative assessments enable direct comparisons of different sample preparation protocols, providing a basis for iterative refinement and optimization of the protocol.

2.3.4. Discussion

Fluorescence microscopy has proven to be an essential tool for analyzing cerebral organoids, offering critical insights into the process of neuronal development. Despite the challenges associated with generating three-dimensional (3D) reconstructions of organoids, which require tissue clearing and advanced imaging techniques [59, 64], the segmentation-based approach presented in this study using epifluorescence microscopy provides an alternative for assessing neuronal development through 2D image slices. Given the limited availability of data, this study focused on developing and evaluating analysis pipelines for nuclei and ventricle segmentation, while also assessing the feasibility of registration-based volumetric reconstruction of cerebral organoids.

The nuclei segmentation model trained on a curated dataset with clear nuclear boundaries demonstrated an Q_{F} score of 0.85, where Q_{F} is a variant of the F1-score used for evaluating cell detection performance. However, the model struggled with noisy and ambiguous data, particularly those with less well-defined nuclear boundaries, suggesting the need for improved image quality through thinner organoid slices or z-stack imaging. Additionally, the proposed method of threshold adaptation for BeadNet predictions offers a promising strategy for semi-supervised nuclei counting, which could prove useful in cases with noisy annotations.

In the case of ventricle segmentation, the use of SegFormer proved significantly more accurate than traditional methods like Otsu's thresholding, achieving Dice scores of 0.9, even surpassing human performance. This improvement highlights the power of modern deep-learning techniques in segmentation tasks. Furthermore, evaluating model certainty can be crucial for optimizing organoid staining and further improving the reliability of ventricle segmentation.

Lastly, the work on registration-based volumetric reconstruction underscores both the challenges and potential solutions for generating 3D models of cerebral organoids. Due to limited data for developing fully automated analysis pipelines, manual pre-registration steps and artifact exclusions were necessary in this study. However, the approach successfully demonstrated the feasibility of reconstructing 3D organoid models from 2D image slices using image registration. Notably, a recent study successfully performed 3D reconstructions of cortical organoids without reporting artifacts during sample preparation [205], raising the question of whether their sample preparation protocols were of higher quality or whether problematic samples were manually excluded from their analysis. Moving forward, automating these preprocessing steps will be essential for achieving a fully automated pipeline. Integrating deep learning models to detect and correct issues such as tissue rotation, mirroring, and folding could significantly enhance the robustness and scalability of organoid reconstruction.

In conclusion, the integration of automated analysis pipelines for epifluorescence microscopy-based segmentation and registration has the potential to transform the analysis of cerebral organoids, while also revealing limitations in current sample preparation protocols. By automating the quantification of neuronal development and enhancing 3D reconstructions, these methods not only provide new insights into neuronal maturation but also offer a pathway for optimizing experimental protocols, ensuring the generation of high-quality data for future studies.

2.4. Cell-Level Analysis

Compared to organoid-level analyses, cell-level analysis provides deeper insights into the composition and characteristics of distinct cell populations within stem cell-derived models. The following section investigates single-cell RNA sequencing (scRNA-Seq) and imaging flow cytometry, focusing on their integration into automated analysis pipelines for brain organoid characterization and iterative protocol optimization.

2.4.1. Single-cell RNA-Seq

> This analysis builds upon data preprocessed by Daniel Hannuschke for his master's thesis [206], which included scRNA-Seq quality control, normalization, feature selection, dimensionality reduction, and exploratory single-cell differential gene expression and differential abundance analyses. Luca Deininger conducted a more targeted investigation focused on identifying SSADH-specific features in cerebral organoids using complementary analysis methods, such as pseudobulk differential gene expression analysis. Additionally, L.D. explored the dataset in the context of optimizing stem cell model protocols. All figures presented here were generated by L.D. Figure 2.31c and 2.35a are similar in [206] due to comparable analysis steps.

In this work, scRNA-Seq data from wild-type and Succinic Semialdehyde Dehydrogenase (SSADH)-deficient cerebral organoids were analyzed to identify disease-specific features, demonstrate the value of cell-level analysis, and explore the potential of scRNA-Seq for optimizing stem cell-derived model protocols.

2.4.1.1. Dataset

Cerebral organoids not previously introduced in earlier sections were cultivated from pluripotent stem cells derived from two SSADH-deficient patients, two parents of SSADH patients, and one Wildtype, with three replicates for each group (Table 2.8).

Organoid batches underwent scRNA-Seq at days 60 and 120 to assess differences in temporal neural development (Table 2.8). While the Parent organoids were included for cell annotation, they were excluded from the differential gene expression analysis and gene set enrichment analysis due to uncertainties regarding their quality, as discussed in Section 2.4.1.4.

2.4.1.2. Methods

Data Preprocessing

As previously mentioned in this section, for detailed information on the data preprocessing steps before cell clustering, please refer to [206]. For all replicates, except for wildtype replicate 2 (day 60), a count matrix could be derived. Wildtype replicate 2 (day 60) was

Table 2.8: Number of technical replicates for single-cell RNA-Seq. Each technical replicate corresponds to one cerebral organoid.

Group	Day 60	Day 120
Patient 1	3	3
Patient 2	3	3
Parent 1	3	3
Parent 2	3	3
Wildtype	3	3

likely lost during sample preparation. Quality control included filtering for cells with at least 200 detected genes, genes present in at least three cells, and thresholds for total read counts and mitochondrial gene content. Specifically, cells with more than 60,000 total reads or more than 5% mitochondrial transcripts were excluded to ensure high-quality input for downstream analysis. The mitochondrial transcript threshold serves as an indicator of cell integrity, as elevated mitochondrial transcript levels are commonly associated with cell stress or apoptosis [207]. Downstream cell clustering was performed using the Louvain [208] algorithm with a resolution of 0.8. Cluster annotation was primarily guided by a cerebral organoid-specific scRNA-Seq tutorial, which included relevant genes for annotating cerebral organoid cell types [209].

Pseudobulk Differential Gene Expression Analysis and Gene Set Enrichment Analysis

The pseudobulk differential gene expression analysis (DGEA) and subsequent gene set enrichment analysis (GSEA) were performed using pairwise comparisons, with one comparison for each combination of cell type and day. For example, Patient organoids were compared to Wildtype organoids for cell type X on day 60. Consequently, each group-wise comparison (e.g., Patients vs. Wildtype) comprises # time points × # cell types individual comparisons.

A higher cell count in cell-wise comparisons could bias the analysis, potentially leading to the erroneous conclusion that certain genes are overexpressed simply because one sample has more cells, thus resulting in a higher number of transcripts (Figure 2.30a). To mitigate this issue, the first step was cell-wise subsampling of the majority class to match the minority class (Figure 2.30b). For instance, if Wildtype had 40 cells for cell type X and Patients had 200 cells, 40 cells were randomly drawn from the Patients to match the Wildtype sample. To mitigate subsampling bias, ten independent subsamples were generated for each cell type (Figure 2.30b). Then, the one exhibiting the highest average Pearson correlation of fold changes to all other subsamples was selected as the representative subsample (A^*, Figure 2.30c). This selected subsample was then used for subsequent pseudobulk DGEA and GSEA (Figure 2.30d).

For DGEA, DESeq2 [83] was employed. Genes with fewer than ten transcripts were excluded before DGEA, which is a standard practice. For GSEA, the gseGO function was applied with 10,000 permutations to assess the statistical significance of the results. The

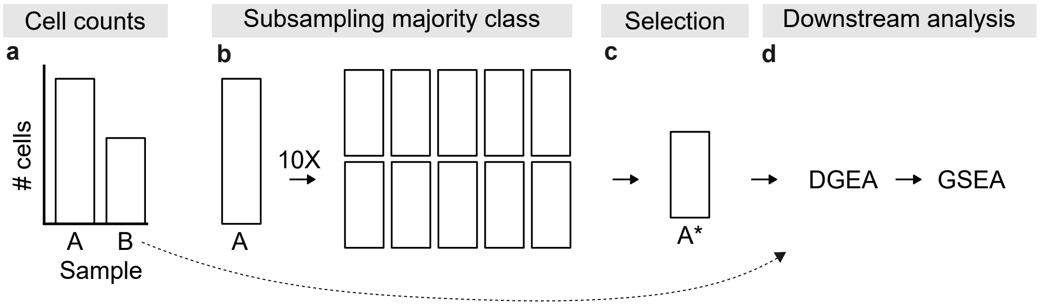


Figure 2.30: Workflow of pseudobulk differential gene expression analysis (DGEA) and gene set enrichment analysis (GSEA). (a) Unequal numbers of cells per condition for a given cell type can introduce bias in the analysis. (b) To address this, cells from the majority group were randomly subsampled to match the number of cells in the minority group. This process was repeated ten times to estimate subsampling variability. (c) The subsample with the highest average Pearson correlation of fold changes to all other subsamples was selected as the representative subsample (A*). (d) Downstream DGEA and GSEA were then performed on A* and B. The entire workflow was applied separately for each cell type.

minimum and maximum gene set sizes considered were 3 and 800, respectively. A p-value cutoff of 0.05 was used to filter genes from the DGEA analysis, and p-value adjustment for multiple testing was performed using the false discovery rate. To focus on neuronal differentiation pathways, the GSEA was restricted to GO gene sets with the parent term *neuron differentiation*.

2.4.1.3. Results

Cell Annotation and Distribution

To explore existing cell populations, a clustered UMAP representation was generated from the gene expression profiles of all 43,007 cells across 36,601 genes, integrating all experimental groups, replicates, and time points (Figure 2.31a). Aggregating the full dataset enhances annotation quality by providing more comprehensive coverage and mitigating potential underrepresentation of certain cell types due to batch effects. Using the Louvain algorithm, 29 cell clusters were identified, each characterized by differential expression of different genes essential for distinguishing cerebral organoid cell populations (Figure 2.31b). The cell annotation revealed that the dataset included most of the expected cell types, such as Neural Progenitor Cells (NPCs) and neurons (Figure 2.31c). Notably, NPCs were further subdivided into several subpopulations based on their anatomical location and cellular characteristics, including cortical NPCs, diencephalon NPCs, oligodendrocytes, and astrocytes. The dataset also contained a variety of neuron subtypes within the cerebral organoid, such as inhibitory neurons from the hindbrain and midbrain, excitatory neurons from the hindbrain, and diencephalic excitatory neurons. Overall, this scRNA-Seq analysis provides a detailed map of the cell populations within the cerebral organoid, highlighting the diversity and complexity of its cellular composition.

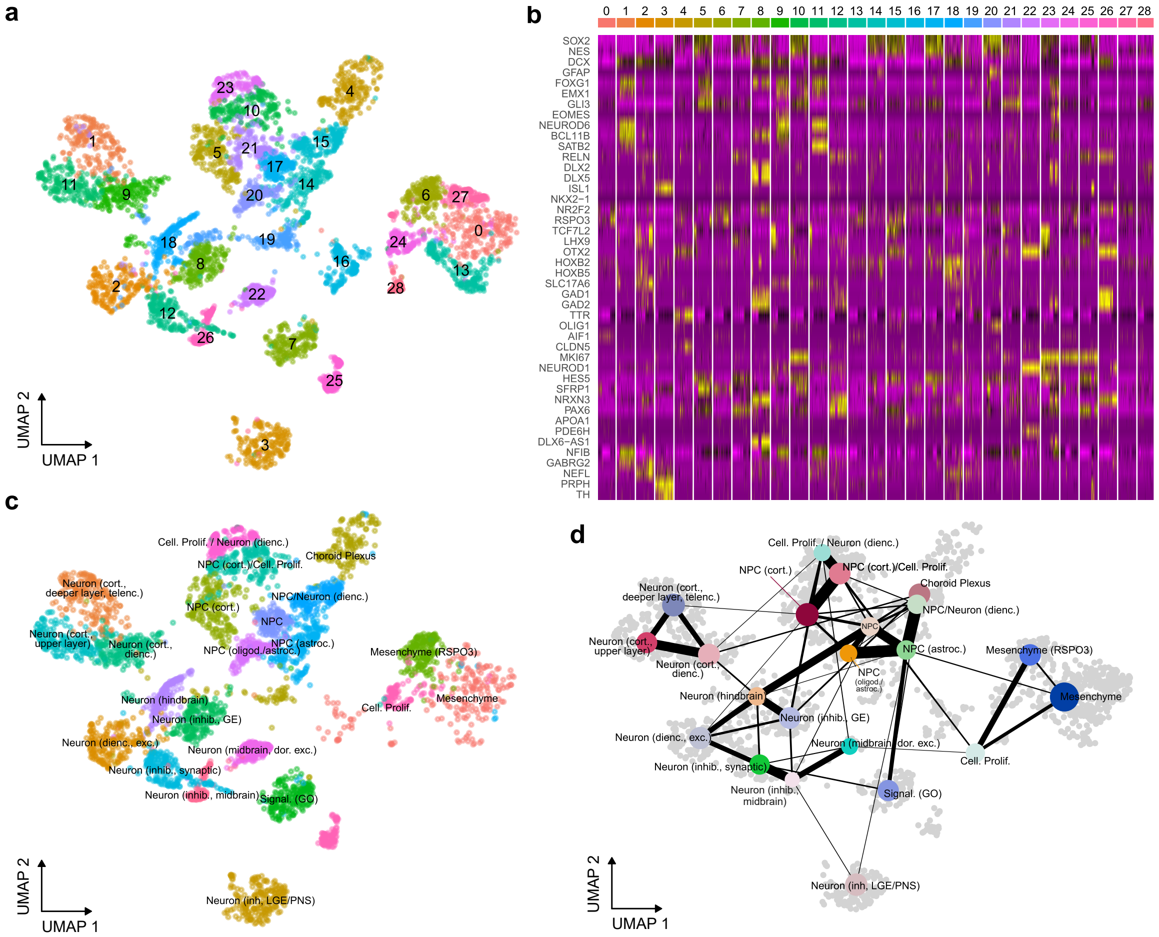


Figure 2.31: scRNA-Seq cell annotation, cell clusters, and cluster similarity. (a) UMAP representation of single-cell gene expression profiles across 43,007 cells and 36,601 genes, integrating all experimental groups, replicates, and time points. Each data point represents a single cell, and the cells are clustered based on their gene expression profiles using the Louvain algorithm. Cells are color-coded by cluster identity, with distinct colors and overlaid cluster numbers indicating the respective clusters. For visualization purposes, 300 cells were randomly sampled from the entire dataset. (b) Heatmap showing cluster-specific expression of cerebral organoid-related genes. (c) UMAP representation from (a) with annotated cell clusters, based on the expression data from (b). Cells are color-coded by cluster identity. For visualization purposes, 300 cells were randomly sampled from the entire dataset. (d) Partition-based Graph Abstraction (PAGA) analysis applied to the UMAP in (c). The edge thickness reflects the strength of transcriptional connectivity between clusters, with thicker edges indicating a stronger connectivity. Nodes are color-coded by cluster identity, but the colors do not correspond to those in (a) or (c). dienc, diencephalon; telenc, telencephalon; inhib, inhibitory; exc, excitatory; oligod, oligodendrocyte; astroc, astrocytes; cort, cortical; signal, signaling.

UMAP representations are non-deterministic, and the relative positioning of cell populations should, to some extent, reflect the similarity between cell clusters. To validate this, PAGA (Partition-based Graph Abstraction) was employed, a tool designed to assess the similarity between cell clusters. A PAGA analysis illustrated that clusters of cells that are spatially close on the UMAP also exhibited higher similarity to each other (Figure 2.31d). For example, NPCs formed tightly connected clusters, as did neurons. This reinforces the idea that the UMAP preserved global relationships between cell clusters. Additionally, PAGA highlights potential differentiation trajectories, tracing the progression from proliferating cells through NPCs to various neuronal subtypes (Figure A.8).

The number of NPCs and neurons at different developmental stages serves as an indicator of neuronal development in cerebral organoids. An increase in NPCs over time was observed in Wildtype organoids, whereas a decreasing trend was detected in Patient organoids (Figure 2.32a). Additionally, Patient organoids already contained more neurons than Wildtype organoids by day 60 (Figure 2.32b). These findings support a previous hypothesis, based on immunohistochemistry data, suggesting premature neuronal differentiation in SSADH patients [52].

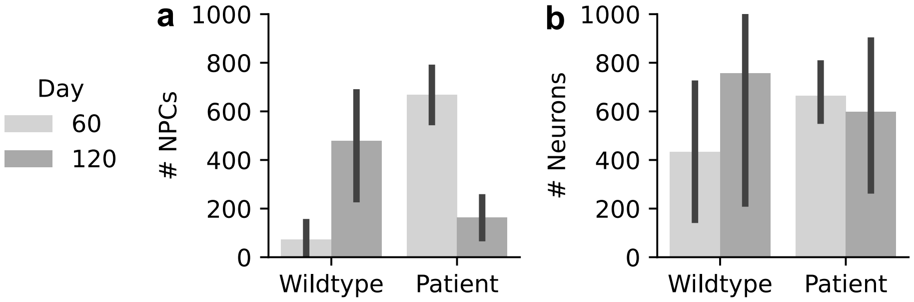


Figure 2.32: Number of NPCs and neurons indicate earlier neuronal development in Patients. The data are derived from cell annotations in Figure 2.31, where NPC and neuron subtypes were aggregated to provide an overview of neuronal development. Error bars indicate 95% confidence intervals based on three technical replicates per group and time point, except for the Wildtype group on Day 60, which included two replicates.

Pseudobulk Differential Gene Expression Analysis and Gene Set Enrichment Analysis

The pseudobulk differential gene expression analysis and gene set enrichment analysis were conducted to investigate SSADH-specific pathways related to neuron differentiation. To this end, Patient and Wildtype organoids were compared at day 60 and day 120 using cell-wise comparisons as explained in Section 2.4.1.2 and Figure 2.30. First, the number of cell types exhibiting enriched neuron differentiation-related pathways for both Wildtype and Patient organoids was compared at each time point. Wildtype organoids showed an increasing number of cell types with enriched pathways related to neuron differentiation over time, while Patient organoids exhibited the opposite trend (Figure 2.33a). For example, midbrain excitatory neurons in Patient organoids at day 60 showed significant enrichment of multiple GO pathways, such as *dendrite development*, *axon development*, *neuron development*, and *neuron projection guidance* (Figure 2.33b), whereas no enriched pathways were observed in Wildtype organoids at this time point. By day 120, however, no enriched pathways were observed in Patient organoids, while Wildtype organoids displayed enrichment in pathways like *neuron development*, *neuron differentiation*, and *axon development* (Figure 2.33c). Additional examples can be found in the Appendix (Figure A.9). This, again, supports the hypothesis of premature neuronal differentiation in SSADH patients.

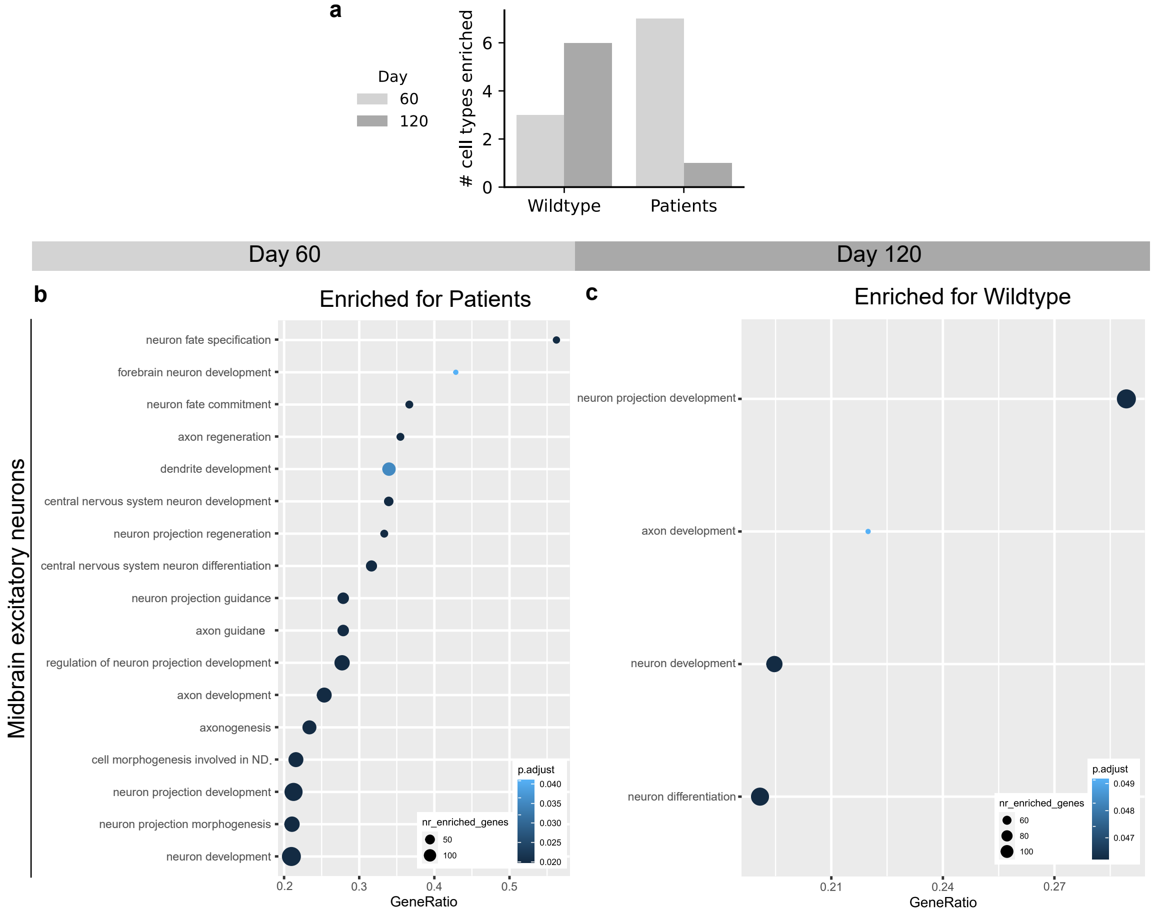


Figure 2.33: Gene set enrichment analysis (GSEA) suggests earlier neuron differentiation in Patients. GSEA was performed on pseudobulk DGEA results comparing Patient and Wildtype organoids, separately for each cell type and time point. (a) The number of cell types with enriched Gene Ontology (GO) terms related to neuron differentiation distinguished between Wildtype and Patients for days 60 and 120. (b,c) Enrichment of neuron differentiation-related GO terms in midbrain excitatory neurons, comparing Patient and Wildtype organoids. The x-axis represents the *GeneRatio*, defined as the proportion of differentially expressed genes in midbrain excitatory neurons that are also present in the given GO term. (b) Day 60. There were no gene sets enriched for Wildtype. (c) Day 120. There were no gene sets enriched for Patients.

2.4.1.4. Protocol Optimization

Presence of Mesenchymal Cells

Recent work has shown that high-quality cerebral organoids typically contain fewer mesenchymal cells, with proportions usually less than 30% mesenchymal cells, whereas an increased mesenchymal cell count is associated with lower-quality organoids [210]. In the present analysis, the fraction of mesenchymal cells was quantified across all samples, revealing lower percentages in both Wildtype and Patient organoids (Figure 2.34). In contrast, Parent organoids exhibited a higher percentage, with mean values of up to 70% for Parent 1 (Figure 2.34). This suggests that the Parent organoids are at the critical threshold for mesenchymal cell abundance. Given the biological uncertainties associated with other transcriptomic effects from lower-quality organoids, as well as the inherent uncertainties regarding the expected biological phenotype for SSADH parents, Parent organoids were excluded from pseudobulk differential gene expression analysis and gene set enrichment analysis.

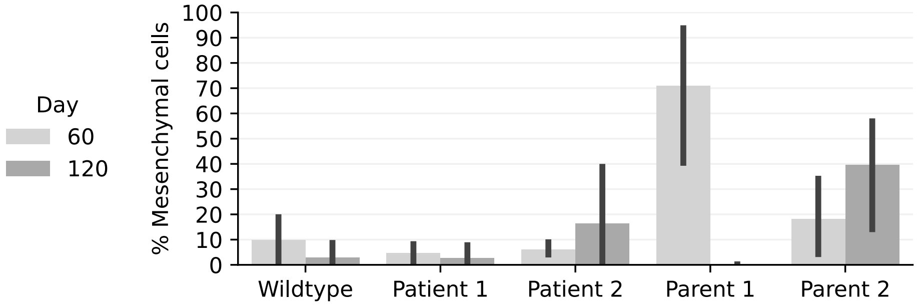


Figure 2.34: Percentage of mesenchymal cells. The percentage of mesenchymal cells is shown per group, aggregated across all replicates. Error bars indicate 95% confidence intervals based on three technical replicates per group and time point, except for the Wildtype group on Day 60, which included two replicates.

Sample Variability

Batch-to-batch variability presents a significant challenge in the study of cerebral organoids. To evaluate the extent of this variability, the correlation of cell counts across replicates was assessed (Figure 2.35a). The results revealed strong correlations in cell counts for replicates of Wildtype on day 60, Parent 1 and 2 on day 120, and Patient 1 on day 60 (Figure 2.35a,b, Table 2.9). Moderate correlations were observed for Parent 2 and Patient 2 on day 60, while weak correlations were found for Patient 1 and Patient 2 on day 120 (Figure 2.35a,b, Table 2.9). Notably, Wildtype on day 120 exhibited a slight negative correlation. Additionally, both Wildtype and Patient organoids exhibited decreasing cell count correlation over time, which may reflect a natural diversification process as organoids mature. While cell counts were moderately or strongly correlated on day 60, these correlations weakened on day 120 (Figure 2.35b, Table 2.9).

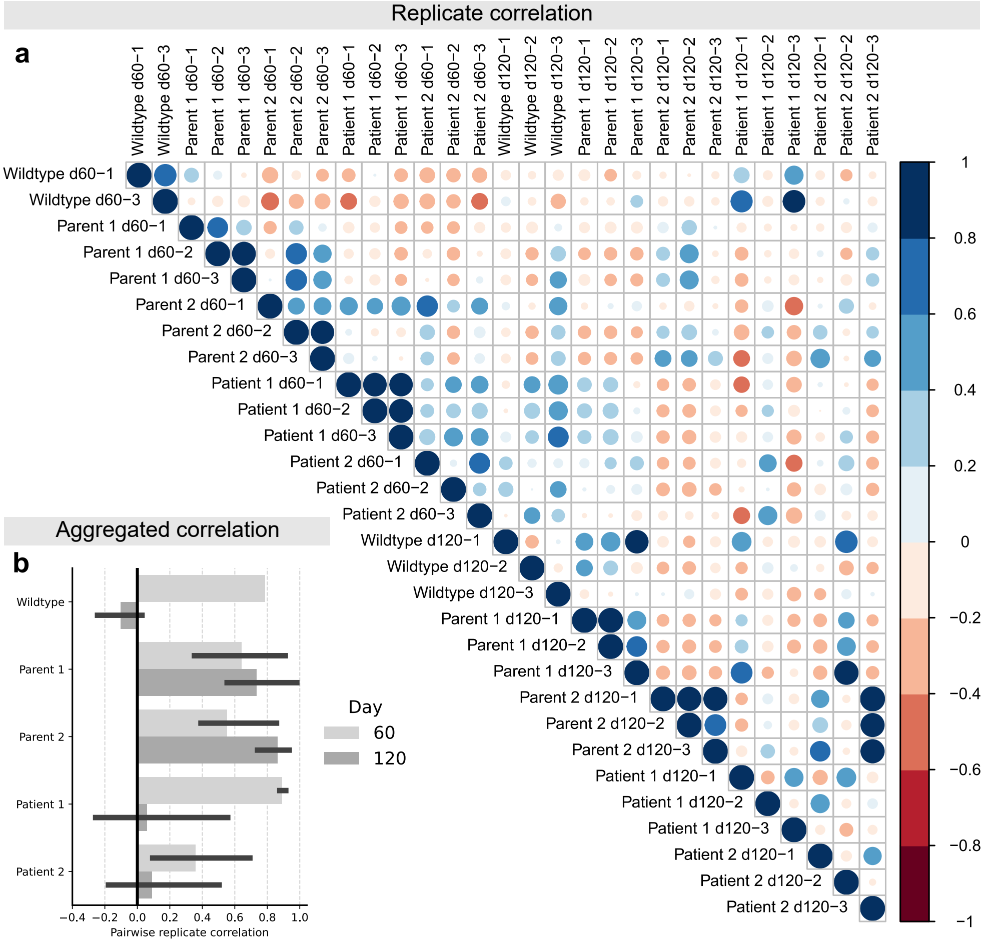


Figure 2.35: Correlation of cell counts across individual replicates reveals organoid variability. Correlations are based on cell counts across 22 annotated cell types. Mesenchymal cells were excluded from the analysis, as their presence is considered an indicator of low-quality organoids [210]. (a) Pairwise correlation of all replicates. (b) Pairwise correlations of replicates from the same sample type. Correlation: Pearson. Error bar, CI 95%.

Overall, this analysis highlights the importance of integrating multiple measures to assess the quality of cerebral organoids using scRNA-Seq. While the Parent organoids exhibit relatively high correlation in cell counts, especially on Day 120, indicating low variability and high reproducibility, the elevated abundance of mesenchymal cells raises concerns about their overall quality (Table 2.9).

Table 2.9: Sample quality based on cell count correlation and mesenchymal cell ratio. Sample quality is assessed based on the mean pairwise cell count correlation across replicates (Q_{CC}), categorized as low ($r < 0.3$), medium ($0.3 \leq r < 0.7$), and high ($r \geq 0.7$); and the mean mesenchymal cell ratio across replicates (Q_{MCR}), considered high quality if $< 10\%$, medium if between 10% and 30%, and low quality if ≥ 30%, based on [210]. It should be noted, however, that lower Q_{CC} at later time points may also reflect increased temporal diversification.

Day	Sample	Q_{CC}	Q_{MCR}
60	Wildtype	High	High
	Patient 1	High	High
	Patient 2	Medium	High
	Parent 1	Medium	Low
	Parent 2	Medium	Medium
120	Wildtype	Low	High
	Patient 1	Low	High
	Patient 2	Low	Medium
	Parent 1	High	High
	Parent 2	High	Low

2.4.1.5. Discussion

This study examined publicly available semi-automated analysis pipelines, including Seurat [84] and DESeq2 [83], for single-cell RNA sequencing (scRNA-Seq) data, customized to identify SSADH-specific features in cerebral organoids. Consistent with prior observations in SSADH patients [52], organoids derived from SSADH patients displayed signs of premature neuronal differentiation, supported by NPC and neuron cell counts as well as gene set enrichment analysis, underscoring the potential of cerebral organoids for disease modeling.

While imaging-based analysis pipelines, such as those for brightfield microscopy, primarily assess organoid size and morphology, scRNA-Seq provides complementary insights by capturing cell-wise gene expression. This enables detailed characterization of cerebral organoid cell populations and the identification of disease-specific features. Additionally, pathway enrichment analysis offers a functional perspective for comparing conditions, further enhancing the biological relevance of scRNA-Seq data.

Despite increasing automation in scRNA-Seq analysis, expert input remains essential. Pipelines like Seurat [84] and Cell Ranger [211] automate preprocessing, clustering, and differential gene expression analysis, yet expert oversight is required for quality control, feature selection, and result interpretation to minimize bias. While automation streamlines tasks such as normalization and gene set enrichment, full automation remains an emerging field, with most pipelines providing only preliminary outputs [212].

Two key metrics are valuable for optimizing cerebral organoid cultivation: the proportion of mesenchymal cells, which serves as an indicator of cerebral organoid quality [210], and the correlation of cell counts, which provides insight into organoid variability (Section 2.4.1.4, Table 2.9). In this dataset, the high proportion of mesenchymal cells in Parent organoids raised concerns about sample quality, leading to their exclusion from differential gene expression and gene set enrichment analyses to reduce confounding effects. Additionally, organoid variability was observed, with correlations in cell counts weakening over time, particularly in Wildtype and Patient organoids at day 120, potentially reflecting natural diversification. These findings underscore the importance of monitoring mesenchymal cell proportions and cell count correlations as key metrics for iteratively refining cerebral organoid cultivation protocols.

Compared to imaging-based modalities, scRNA-Seq enables semi-automated analysis of cell populations, offering deeper biological insights. Advancing toward fully automated scRNA-Seq could further enhance disease modeling and stem cell research [212]. While multimodal characterization, such as combining brightfield imaging, MRI, and scRNA-Seq, would offer a more comprehensive view of organoid structure and function, it remains challenging due to the destructive nature of RNA sequencing. One practical solution is to apply scRNA-Seq as a terminal readout (see e.g. Figure 2.2) or to perform it on separate organoid batches. However, the latter approach introduces batch-to-batch variability, potentially complicating interpretation. Hybrid approaches such as spatial transcriptomics offer promising alternatives by combining transcriptomic profiling with spatial resolution, enabling molecular analysis in a spatial context [213, 214].

In conclusion, the analysis confirmed premature neuronal differentiation in SSADH patient-derived cerebral organoids. These findings underscore the value of cell-level scRNA-Seq for in-depth disease understanding. However, optimizing cerebral organoid culture protocols remains essential to enhance sample quality and reduce variability for more robust disease modeling, with verified metrics such as mesenchymal cell proportions and cell count correlations serving as key indicators for refinement.

2.4.2. Imaging Flow Cytometry

> Parts of this section are based on: D. Vonficht, L. Jopp-Saile, S. Yousefian, et al. "Ultra-high scale cytometry-based cellular interaction mapping". Accepted at Nature Methods. 2025. [215]. The segmentation and analysis pipeline was jointly conceptualized by Luca Deininger and Angelo Yamachui Sitcheu, with A.Y.S. focusing primarily on cell representation in the latent space and L.D. focusing on cell segmentation and cell morphology quantification.

ScRNA-Seq enables detailed characterization of cellular composition by capturing gene expression profiles at single-cell resolution (Section 2.4.1). In contrast, imaging flow cytometry (IFC) provides quantitative insights into cell morphology by combining microscopy with flow cytometry.

Here, key components of an automated analysis pipeline for imaging flow cytometry are outlined to enable a preliminary assessment of its potential for analyzing single cells in brain organoids. The aim is to illustrate how such a pipeline could support automated quantification of cell distributions and morphological features.

2.4.2.1. Dataset

Due to the unavailability of brain organoid imaging flow cytometry data at the time of thesis submission, the biological samples examined in this study consist of human blood cells, including various immune cells such as T-cells and B-cells, which were analyzed using imaging flow cytometry. The dataset includes two groups: a control group with no treatment and a condition group treated with blinatumomab, a drug designed to enhance cell-cell interactions. In total, the control group included 91,668 images, while the condition group comprised 54,179 images. Each sample consisted of multi-channel imaging data acquired across ten channels, including two brightfield and eight fluorescence channels (Figure 2.36). The fluorescence channels captured multiple cell surface and intracellular markers, including CD19-PE-Cy7 (channel 6) to identify B-cells and CD3-BV605 (channel 9) to identify T-cells (Figure 2.36).

T- and B-cell markers are generally mutually exclusive, as lineage-specifying transcription factors enforce distinct gene expression programs during lymphocyte development [216]. Therefore, applying intensity-based thresholding on both markers using the mean fluorescence intensity per cell should, in principle, allow for clear identification of the underlying cell type—either T-cell (T^+B^-) or B-cell (T^-B^+). However, technical issues such as staining artifacts may occasionally result in double-positive signals (T^+B^+), complicating interpretation.

Although this study was based on peripheral blood samples, many of the analytical principles are expected to be transferable to brain organoids, especially given recent applications of flow cytometry in this context [217, 218].

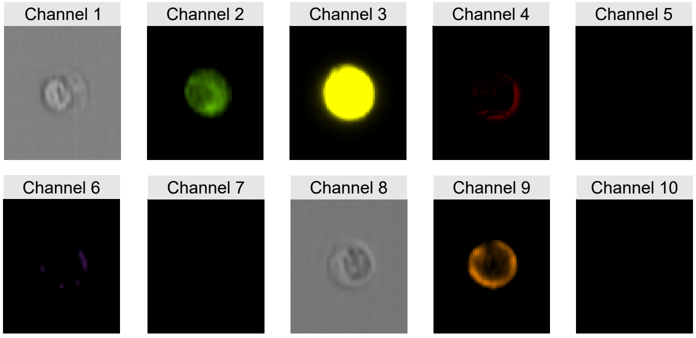


Figure 2.36: Exemplary multi-channel image from imaging flow cytometry. Depicted is a single cell imaged across ten channels, including two brightfield channels (channels 1 and 8) and nine fluorescence channels (remaining channels). The fluorescence channels visualize various surface and intracellular markers, such as CD19-PE-Cy7 in channel 6 for B-cell detection and CD3-BV605 in channel 9 for identifying T-cells.

2.4.2.2. Methods

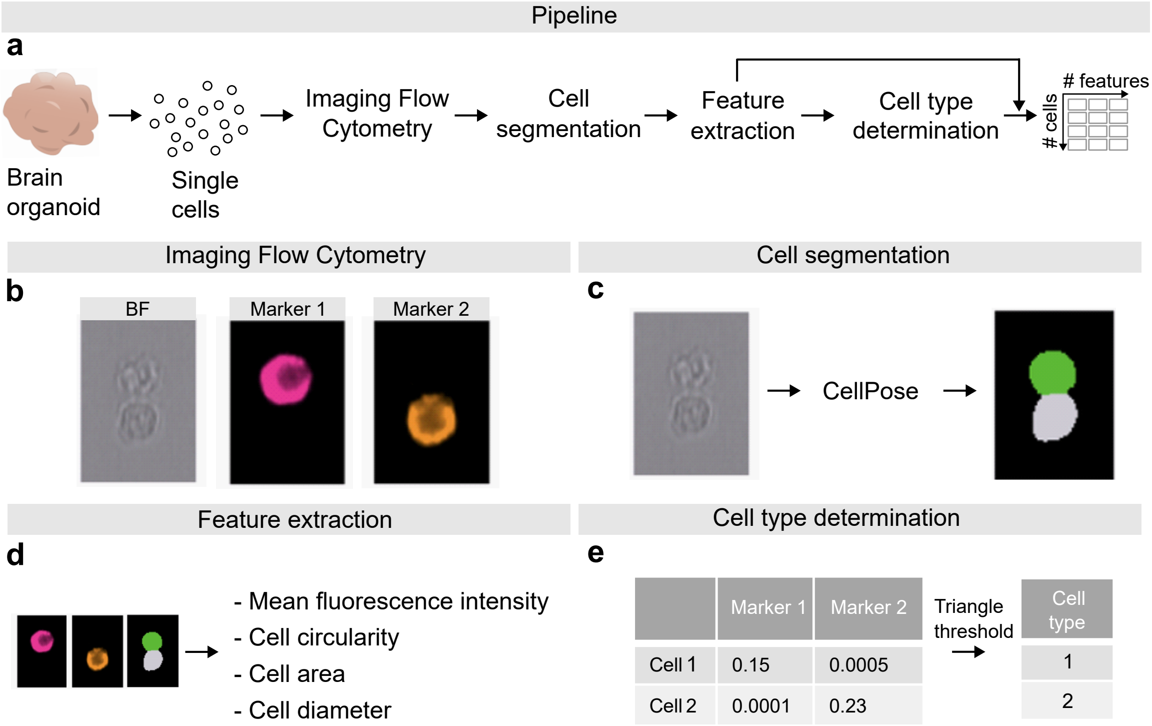
For cell segmentation from brightfield images, the cyto2 model from the Python package CellPose [109] was used with a manually tuned cell diameter of 20 pixels to achieve optimal segmentation results. Although CellPose offers automatic cell size estimation, it was not used due to limited accuracy at the time of this study. Other methods, such as Mesmer [219], infer cell size directly from image features. However, Mesmer is specifically optimized for tissue imaging, whereas CellPose is designed as a general-purpose model applicable to a broader range of microscopy data. Given these considerations and the nature of the data, CellPose was selected. Circularity was calculated based on cell area and perimeter (Equation 2.18), with perimeter values obtained from scikit-image.

$$Circularity = \frac{4 \cdot \pi \cdot \text{Area}}{\text{Perimeter}^2} \tag{2.18}$$

Mean fluorescence intensities for each channel were computed by masking the images of the respective fluorescence channel with the cell segmentation mask. Cell type determination was based on the triangle threshold [220] applied to the T-cell and B-cell channel mean fluorescence intensities to distinguish between T^+ and T^- cells, and B^+ and B^- cells. The triangle threshold was selected because it showed the smallest difference compared to expert-annotated thresholds on small data subsets. T^+B^- cells were classified as T-cells, T^-B^+ cells as B-cells, and T^-B^- cells as *Other*. For T^+B^+ cells, class membership was determined by comparing the mean intensity of the T-cell and B-cell channels, with the marker exhibiting the higher intensity dictating the class assignment. This approach assumes that the dominant signal reflects the true underlying cell identity in cases of ambiguous or artifactual double-positivity.

2.4.2.3. Results

The proposed pipeline begins with the dissociation of brain organoids into single cells, followed by IFC (Figure 2.37a,b). Brightfield images are then used for cell segmentation via CellPose, a state-of-the-art model for cell segmentation (Figure 2.37c). Cell morphology features, such as cell area and circularity, are subsequently extracted, along with the mean fluorescence intensity for each cell (Figure 2.37d). In the next step, the cell type is determined through marker-specific thresholding (Figure 2.37e, details in Section 2.4.2.2). This process results in a feature table of size # cells × # features, which can then be used to analyze cell populations and morphological characteristics.



	Marker 1	Marker 2
Cell 1	0.15	0.0005
Cell 2	0.0001	0.23

Cell type
1
2

Figure 2.37: Concept for an automated analysis pipeline for brain organoids based on imaging flow cytometry. The data shown in (b) is from blood cells, as brain organoid imaging flow cytometry data were unavailable at the time of thesis submission. However, the pipeline is expected to be transferable to brain organoids, supported by existing studies that have successfully applied imaging flow cytometry to progenitor cells of brain organoids [81].

The analysis of cell distributions revealed similar proportions of cell types between the control and condition (Figure 2.38). However, a comparison of cell morphology highlighted significant differences in cell diameter, cell area, and cell circularity across all cell types. For instance, cells in the condition exhibited a significantly larger diameter, particularly for T-cells and B-cells, compared to the control (Figure 2.38). This shows that the pipeline enables single-cell morphological readouts and facilitates direct comparative analyses between experimental conditions.

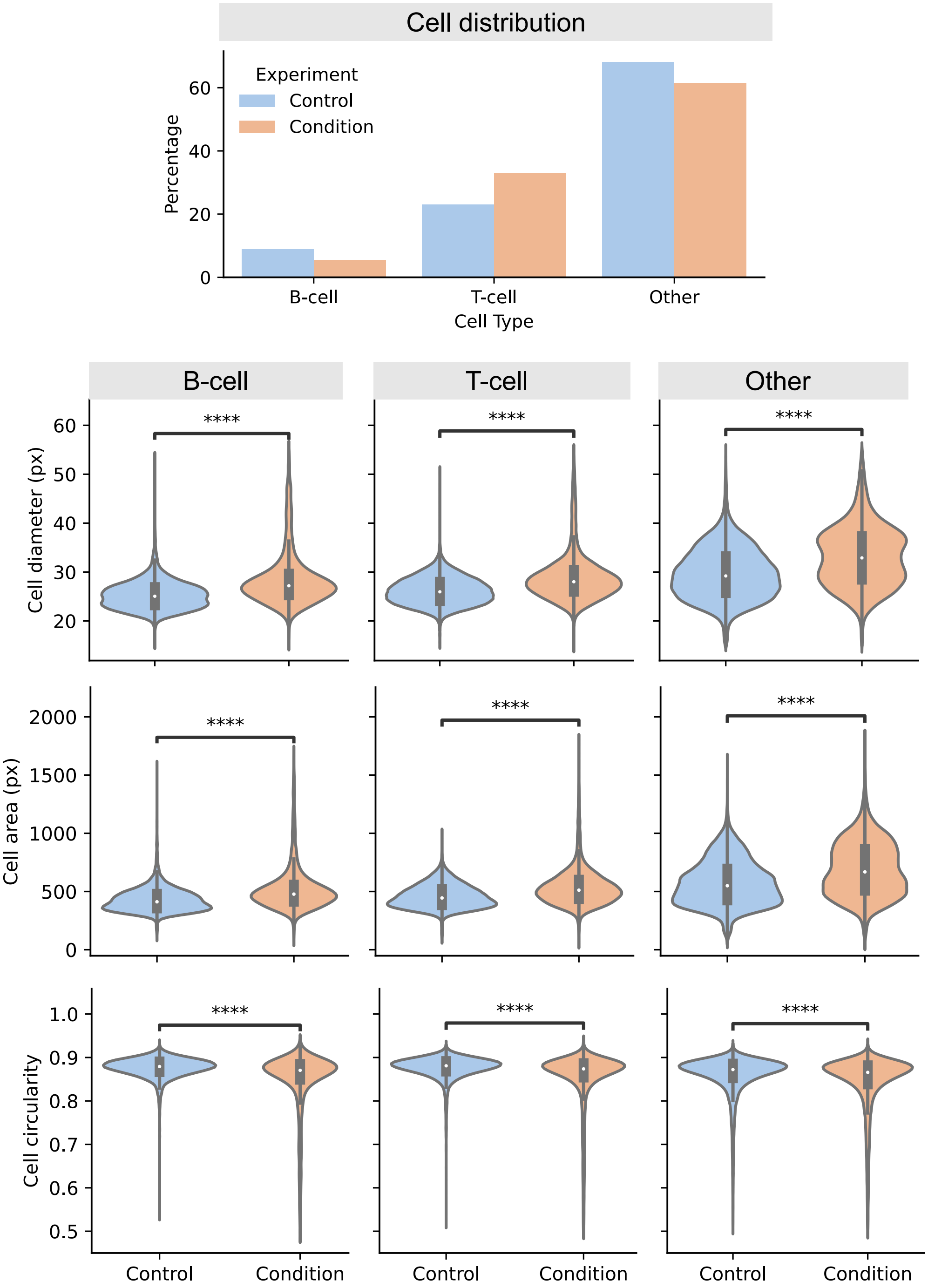


Figure 2.38: Automated analysis pipeline based on imaging flow cytometry data enables comparison of cell distribution and morphology. Comparison between control and condition for two human blood samples, as brain organoid imaging flow cytometry data was unavailable at the time of thesis submission. P-value: Mann-Whitney U test, **** $p < 0.0001$.

2.4.2.4. Protocol Optimization

The correlation of cell distributions used for scRNA-Seq (Figure 2.35) can be similarly applied to imaging flow cytometry data, where a high correlation of cell types reflects lower variability and higher reproducibility. Due to the lack of multiple replicates, this analysis was not performed here.

2.4.2.5. Discussion

This study introduces a concept for an automated IFC pipeline for analyzing single cells from brain organoids. The pipeline involves cell segmentation, feature extraction, and threshold-based cell type determination. The output is a comprehensive cell feature table that enables analysis of cell distributions and morphology, providing clear insights into cellular characteristics.

While imaging-based analysis pipelines primarily evaluate organoid morphology, scRNA-Seq offers complementary insights by capturing cell-wise gene expression. In contrast, IFC provides both cell distribution data, similar to scRNA-Seq, and additional morphological information that scRNA-Seq cannot capture. These advantages position IFC as a valuable modality for bridging the gap between molecular and morphological analyses in brain organoid research.

Due to the unavailability of organoid data, the analysis was conducted on human blood cells. This analysis showed similar cell distributions between control and condition groups but revealed significant morphological differences, particularly in cell diameter, area, and circularity. Treated T-cells and B-cells exhibited larger diameters, indicating morphological changes induced by the treatment. These findings demonstrate IFC's capability to automatically detect subtle changes using the proposed pipeline.

A limitation of this study is the uncertainty regarding the pipeline's applicability to brain organoids. The complex morphology of brain organoid cells, particularly neurons with branching, tree-like structures, may pose challenges for imaging and accurate segmentation. However, previous studies have successfully applied flow cytometry to stem cells [221] and brain organoids [217, 218], and imaging flow cytometry has been used to study neural progenitor cells [81], which are also present in brain organoids. These findings suggest that a similar approach may be feasible for analyzing dissociated brain organoid cells.

Overall, this pipeline is a tool for high-throughput, single-cell analysis with potential for organoid research. This study demonstrates its effective application in determining cell populations and morphology. Moreover, the correlation of cell counts among technical replicates, as used in scRNA-Seq, can serve as a valuable metric for iterative protocol optimization. Future experiments should explore the feasibility of applying IFC to dissociated cells of brain organoids.

3. Analysis Pipelines for ETiX-Embryos

> Parts of this chapter are based on: P. Caldarelli, L. Deininger, S. Zhao, et al. "AI-based approach to dissect the variability of mouse stem cell-derived embryo models". In: *Nature Communications*, Vol. 16, No. 1 (Feb. 2025). ISSN: 2041-1723. DOI: `10.1038/s41467-025-56908-5`. [40]. P.C. was responsible for ETiX-embryo cultivation and data annotation, while L.D. contributed through data processing, analysis, and development of deep learning models.

While previous analysis pipelines have addressed phenotyping and quality control in organ-level stem cell-derived models, comparable challenges arise in stem cell models of whole-organism development. To extend automated analysis to these more complex systems, existing pipeline elements must be adapted to support quality monitoring and protocol optimization.

Here, deep learning-based analysis pipelines were developed for ETiX-embryo classification at the advanced stage around four days post-cell-seeding (Section 3.3.1), where selection of high-quality samples is traditionally performed by human experts. Additionally, early-stage classification was explored to evaluate its accuracy for even earlier ETiX-embryo selection and to identify embryo features predictive of future successful development (Section 3.3.2). AI-based methods were also applied to track the development of individual ETiX-embryos over time (Section 3.3.3). Finally, insights from the pipeline and resulting models informed protocol adjustments, resulting in an increased proportion of successfully developed ETiX-embryos (Section 3.5.2).

3.1. Data and Data Preprocessing

Data

A live-imaging platform was employed to monitor the development of ETiX-embryos over the first 90 hours using confocal microscopy, capturing brightfield and fluorescence images (Figure 3.1). To facilitate cell-type- and substructure-specific tracking during development, fluorescent markers were applied to the three starting cell types: ESC cells (represented in red throughout the text), ESC-iGata4 cells (in green), and TSC cells (in blue). The images are captured as mosaics, consisting of 25 (5 × 5) tiles (Figure 3.1). Acquisition was performed every 35.26 minutes, capturing z-stacks with a spacing of 4 μm. At each time point, the imaging sequence followed a fixed order: for each of the 25 mosaic tiles, a full z-stack was acquired, beginning with brightfield and followed by the three fluorescence channels. This process was repeated tile by tile until the full mosaic was completed.

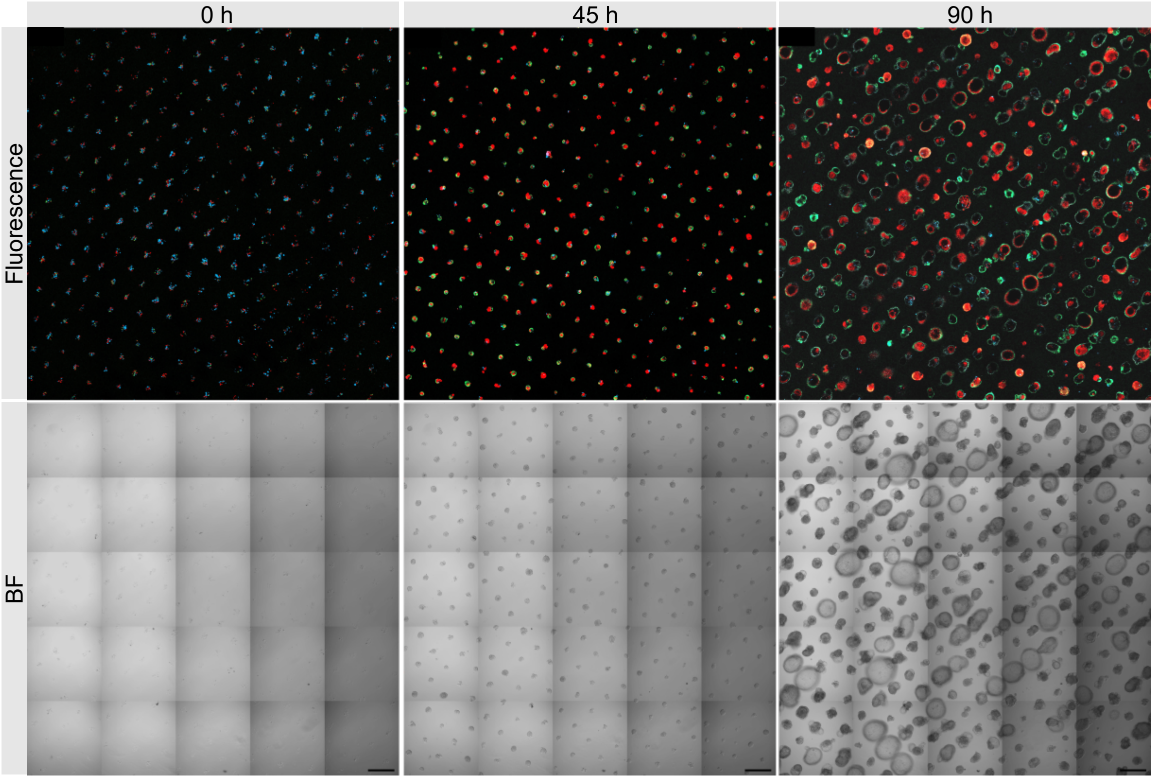


Figure 3.1: ETiX-embryo live-imaging. Representative time points of live imaging movies displaying around 320 ETiX-embryos at 0, 45, and 90 hours after seeding in the Agarwell. Scale bar: 1,000 μm. Adapted from Fig. 1 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with added brightfield images and modified layout.

Three runs (datasets A, B, and C) were recorded, containing 306 (A), 293 (B), and 301 (C) ETiX-embryos, respectively. Together, these runs resulted in a combined dataset of 900 ETiX-embryos. Unless stated otherwise, references in the text refer to the union of these three datasets. Based on insights gained during pipeline optimization, two additional datasets were subsequently generated, featuring increased initial cell counts across all cell types: Data_{2X} with a twofold increase ($n = 306$), and Data_{3X} with a threefold increase ($n = 276$).

To account for the different spatial scales present in the dataset, coordinate sets were defined separately for each image level: tile, mosaic, and extracted embryo.

- Let $X_{\text{tile}}, Y_{\text{tile}}$ denote the x and y coordinate sets of a single tile, with cardinalities $|X_{\text{tile}}| = |Y_{\text{tile}}| = 768$.
- Let $X_{\text{mos}}, Y_{\text{mos}}$ denote the coordinate sets for the full mosaic image, formed by a 5×5 grid of tiles with $|X_{\text{mos}}| = |Y_{\text{mos}}| = 3840$.

- Let X_{emb}, Y_{emb} denote the coordinate sets of an extracted embryo image (Section 3.1), where $|X_{\text{emb}}| = |Y_{\text{emb}}| = 288$.
- Let Z, T, and C denote the sets of z-planes (depth), time points, and imaging channels, respectively, with $|Z| = 40$, $|T| = 153$, and $|C| = 4$. The four channels correspond to one brightfield and three fluorescence channels.

At the beginning of this thesis, no comparable imaging datasets of similar scale for ETiX-embryos were publicly available, underscoring the value of the data generated in this study. All images preprocessed for deep learning generated for this work are publicly available on Zenodo (`https://zenodo.org/records/14605093`). The code to reproduce the results is publicly available on GitHub (`https://github.com/deiluca/StembryoNet`).

Preprocessing

The brightfield mosaic tiles exhibited varying brightness levels (Figure 3.1). To address this brightness variation, the brightfield mosaic tiles were corrected by normalizing their intensities within the same z-plane and time point. Let $\mathcal{M}$ denote the set of mosaic tiles m. For each z-position $z \in Z$ and time point $t \in T$, each tile was normalized by the mean intensity of all 25 tiles in the corresponding z-plane at time point t:

$$m(z,t)_{norm} = m(z,t) \cdot \frac{\text{Mean Intensity } (\mathcal{M}, z, t)}{\text{Mean Intensity } (\mathcal{M} = \{m\}, z, t)} \tag{3.1}$$

$$\text{Mean Intensity } (\mathcal{M}, z, t) = \frac{1}{|\mathcal{M}| \cdot |X_{\text{tile}}| \cdot |Y_{\text{tile}}|} \sum_{m \in \mathcal{M}} \sum_{x \in X_{\text{tile}}} \sum_{y \in Y_{\text{tile}}} m(x, y, z, t) \tag{3.2}$$

As an alternative, shading correction with the BaSiC tool [222] was tested but produced comparable results. Therefore, the simpler mean-based normalization was retained in this study.

Extraction of Individual ETiX-Embryo Images

To generate a single image file per ETiX-embryo, manual segmentation was performed at the final time point. This manual step was necessary because overlapping ETiX-embryos in some cases made automated segmentation unreliable. Automation of this step is discussed in Section 3.5.1. The center point of each segmented ETiX-embryo was then used to extract individual images with dimensions $|X_{\text{emb}}| \times |Y_{\text{emb}}| \times |Z| \times |C| \times |T|$. This approach relies on the implicit assumption that embryo positions remain consistent over time, which is valid due to minimal embryo movement. In cases where ETiX-embryos were growing within the imaging frame of neighboring ETiX-embryos, any overlapping regions were masked. As a result, parts of the image corresponding to neighboring ETiX-embryos were replaced with a value of 120 for brightfield and 0 for fluorescence. ETiX-embryos that extended beyond the imaging border were excluded. Across all three runs, a total of 900 individual ETiX-embryos were obtained.

Data Types for Deep Learning

The extracted images of individual ETiX-embryos are five-dimensional: x, y, z, channel, and time. However, deep-learning models for image classification are primarily designed to process three-dimensional inputs, typically including two spatial dimensions and the channel dimension. To accommodate this constraint, various data input strategies were investigated to compress the z-dimension and incorporate different channel combinations, optimizing feasibility for deep learning training:

- Brightfield in-focus images: These were obtained by selecting the z-plane where the brightfield images appeared most sharply focused. The in-focus plane was determined using the z-plane with the maximum Laplacian variance, a commonly applied method for automatic focus detection [223] (Figure 3.2a).
- Fluorescence in-focus images: These images were extracted from the same z-plane where the brightfield image was in focus, identified based on the maximum Laplacian variance. This approach was chosen as relevant structures in fluorescence images were observed to align with the brightfield in-focus plane (Figure 3.2b).
- Fluorescence z-sum projection images: These were generated by summing pixel intensities along the z-axis, effectively integrating information across all z-slices in the stack (Figure 3.2c).
- Four-channel images: A four-channel representation consisting of brightfield and fluorescence in-focus images ($|X_{\text{emb}}| \times |Y_{\text{emb}}| \times |C|$).

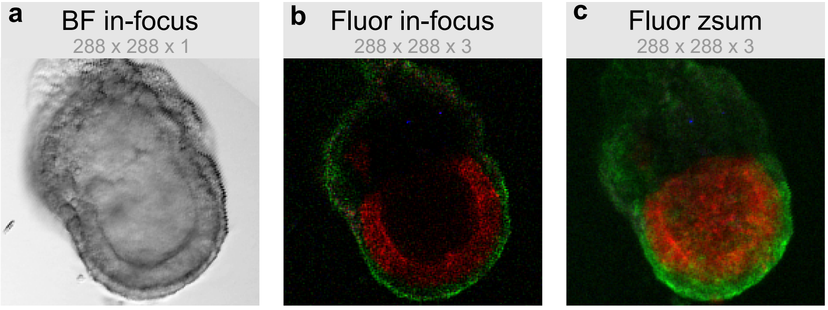


Figure 3.2: Data types explored for deep learning training. (a) Brightfield in-focus image, (b) fluorescence in-focus image, and (c) fluorescence z-sum projection. The same ETiX-embryo is shown in all subplots.

Annotation

ETiX-embryos were annotated based on images from the final 25 hours of time-lapse (Figure 3.3a). ETiX-embryos were classified by an embryologist as *normal* if they displayed a cylindrical shape with distinct cellular compartments derived from TSCs and ESCs, enveloped by a monolayer of ESC-iGata4 cells. Of the ETiX-embryos analyzed in three independent experiments, 23% met the criteria for normal development throughout the observation period (Figure 3.3a,b), while the remaining 77% were classified as *abnormal*, showing structural and developmental abnormalities (Figure 3.3a,c).

The time-lapse studies revealed asynchrony in ETiX-embryo development. To account for this, an end time point was annotated for each normal ETiX-embryo at a comparable developmental stage, ranging from 65 to 90 hours post-cell-seeding (Figure 3.3d-f). For abnormal embryos, direct synchronization was not feasible. Instead, their synchronized time points were sampled using a normal distribution with the same mean and standard deviation as those of normal ETiX-embryos (Figure 3.3d). The resulting subset of images is hereafter referred to as the synchronized dataset, was then used for training the deep-learning model.

For a multi-annotator comparison, three embryologists (A1, A2, and A3) created annotations. While all annotators were included for the multi-annotator comparison, annotations from A1 were used for all other analyses. To benchmark annotation speed, the time required by A2 and A3 to annotate the entire dataset of 900 ETiX-embryos was recorded.

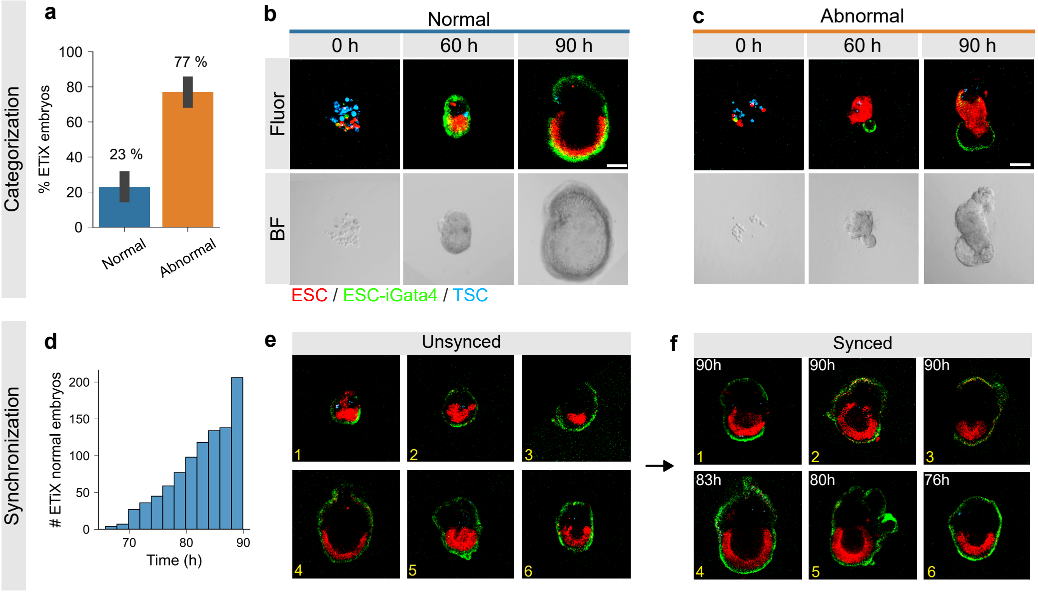


Figure 3.3: Categorization and synchronization of ETiX-embryos. (a) Bar chart displaying the percentage of ETiX-embryos identified as normal or abnormal, with error bars representing variability across datasets (95% confidence interval). (b,c) Sequential fluorescence and bright field images showcasing normal (b) and abnormal (c) ETiX-embryo development. (d) Cumulative histogram of expert-selected (*synchronized*) time points for normal ETiX-embryos. (e) Examples illustrate that normal ETiX-embryos, despite being cultivated simultaneously, displayed different developmental stages at 70 hours. (f) The same embryos as in (e), labeled with yellow numbers in the bottom-left corner of each image, were annotated by an expert embryologist, indicating comparable developmental stages at different time points. Scale bars: 100 μm. Adapted from Fig. 1 and Fig. S2 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0.

3.2. Methods

3.2.1. Advanced-Stage Classification

StembryoNet

For deep-learning-based ETiX-embryo classification at 90 hours post-cell-seeding, StembryoNet, a novel model based on the ResNet18 [95] architecture, is introduced (Figure 3.4). It is specifically designed to be trained on synchronized data while enabling predictions on unsynchronized data. The model retains the five sequential convolutional layers and global average pooling of ResNet18 but replaces the original 1,000-neuron fully connected layer with a single-neuron layer for binary ETiX-embryo classification. Additionally, the softmax function is replaced with a sigmoidal activation function to compute class probabilities. ResNet was chosen for its rapid and reliable convergence in image classification [172]. Furthermore, among various deep learning architectures, ResNet exhibited the smallest standard deviation, indicating greater stability and robustness (Figure 3.7a).

A key feature of StembryoNet is its ability to predict outcomes from unsynchronized data by processing consecutive embryo images from the last 25 hours of development (Figure 3.4, Table 3.1). For each time point, the model outputs a probability that the ETiX-embryo is normal. These probabilities are concatenated across time, and the maximum value is thresholded by θ to derive the final classification.

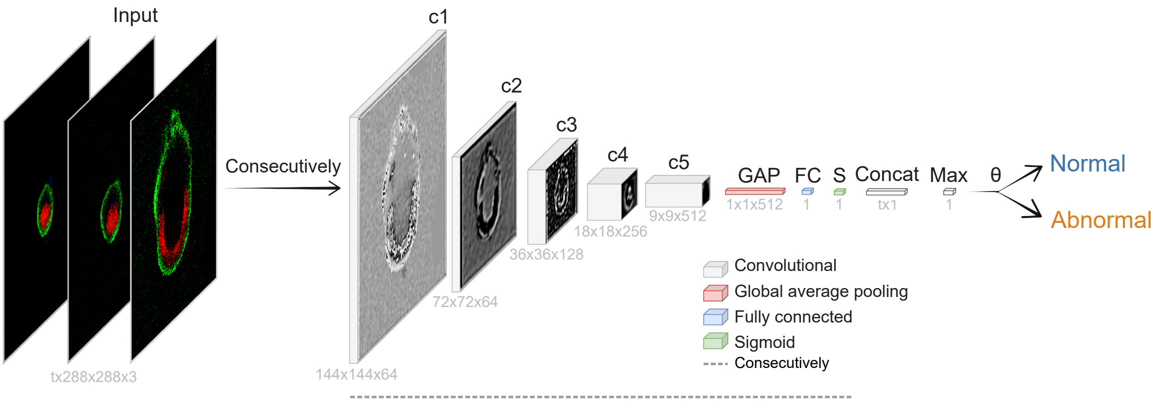


Figure 3.4: StembryoNet architecture. Consecutive images from the final 25 hours (65 to 90 hours post-seeding) of each embryo development are input into StembryoNet for inference. StembryoNet comprises five convolutional layers, followed by a global average pooling (GAP) layer, a fully connected (FC) layer, and a sigmoidal activation function (S). The model predicts the probability of the normal class at each time point. These probabilities are concatenated, and the maximum probability is thresholded by a parameter θ to determine the class. Adapted from Fig. 2 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout and colors.

The StembryoNet backbone was trained on synchronized data, where an expert embryologist annotated the time point of similar development for each embryo (Figure 3.3e,f, Table 3.1). Training was performed on these synchronized time points for 200 epochs with a batch size of 16, using the Adam optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 0.0001) and a binary cross-entropy loss weighted by inverse class frequencies. On-the-fly image augmentations comprised random rotation ($0 - 360$ degrees) and color jittering (using PyTorch's ColorJitter), which involved adjusting the saturation (factor = 0.1) and contrast (factor = 0.1). The images were normalized according to ImageNet default normalization. Model training was implemented with PyTorch Lightning (version 2.0.4).

Table 3.1: Number of time points each model used for training, validation, and testing. The 43 time points correspond to the final 25 hours of the ETiX-embryo video sequences, which span a total of 90 hours. For unsynchronized data, the single time point corresponds to the 90-hour mark, while for synchronized data, it corresponds to the synchronized time point. * Based on synchronized data; † For model selection; ‡ For determination of best threshold (θ). If there is no indicator, the model uses unsynchronized data. ResNeXt, MobileNet, GoogleNet, and DenseNet were configured identically to ResNet.

Model	**Training**	**Validation**	**Testing**
ResNet	1	1	1
MViT	43	43	43
StembryoNet	1^*	$1^{*,\dagger}$, $43^\ddagger$	43

The threshold θ is chosen to maximize the F1-score on the unsynchronized validation set (Table 3.1, Algorithm 1), balancing the trade-off between precision and recall for identifying normal embryos. To compute θ, the maximum probability over the final 25 hours is extracted from the StembryoNet backbone for each embryo in the validation set (max_predictions). The unique values in max_predictions serve as candidate thresholds. For each of these candidate thresholds, binary predictions (binary_predictions) are generated by classifying embryos as normal (1) or abnormal (0), depending on whether their maximum probability exceeds the threshold. The F1-score is then calculated using the ground truth labels and the corresponding binary_predictions. The final threshold θ is defined as the candidate threshold that yields the highest F1-score.

To compare with StembryoNet, several models were trained using the same training procedure as the StembryoNet backbone. These included ResNet18, MobileNet [94], ResNeXt [98], GoogleNet [97], and DenseNet [96], all trained on images captured at 90 hours.

For additional validation, StembryoNet's accuracy and efficiency were compared against three embryologists (A1, A2, and A3). StembryoNet was trained separately using each embryologist's annotation. The corresponding models are referred to as $\text{StembryoNet}_{\text{A1}}$ (SN_{A1}), $\text{StembryoNet}_{\text{A2}}$ (SN_{A2}), and $\text{StembryoNet}_{\text{A3}}$ (SN_{A3}). In this multi-annotator comparison, A1–A3 were included, while annotations from A1 were used for all other analyses, and StembryoNet refers to the model trained on A1's labels. To benchmark annotation speed, the time required by StembryoNet to predict labels for the entire dataset of 900 ETiX-embryos across five repeated 5-fold cross-validation runs was measured, using an NVIDIA A100-40 GPU.

Algorithm 1 Get optimal threshold (θ)

```
Input: max_predictions    // Maximum StembryoNet backbone prediction over the final 25 hours for each embryo
       targets            // Ground truth labels
Output: $\theta$
Initialize $\theta \leftarrow$ None
Initialize $F1_{best} \leftarrow 0$
thresholds ← Unique(max_predictions)
for each threshold ∈ thresholds do
    binary_predictions ← (max_predictions ≥ threshold)
    $F1_{current} \leftarrow$ F1_SCORE(targets, binary_predictions)
    if $F1_{current} > F1_{best}$ then
        $F1_{best} \leftarrow F1_{current}$
        $\theta \leftarrow$ threshold
    end if
end for
Return $\theta$
```

Multiscale Vision Transformer

For the Multiscale Vision Transformer (MViT) [102], the MViT-B architecture pretrained on Kinetics-400 was used, with a frame length of 16 and a sample rate of 4. Training was performed on ETiX-embryo videos spanning 65–90 hours (Table 3.1) for 20 epochs, using a batch size of 1 and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay = 0.0001). The PyTorch automatic learning rate finder was applied before each training to determine an optimal initial learning rate. To effectively utilize the pretrained weights, 16 frames were uniformly sampled from the 65–90 hour time window for both training and testing. MViT training was conducted using PyTorch Lightning (version 2.0.4) and PyTorchVideo (version 0.1.5).

Self-supervised Learning

To explore the potential of self-supervised learning for ETiX-embryo classification, a DINO [116] model was trained on fluorescence images using the ViT_{base} architecture for 200 epochs with a batch size of 32. Default parameter values from the DINO GitHub repository [116] were used for all remaining settings. For hierarchical clustering of DINO embeddings, average linkage clustering based on Euclidean distance was applied. To assess the effects of random hierarchical clustering, ETiX-embryo labels were shuffled, providing a baseline comparison for evaluating DINO clustering performance. For ETiX-embryo classification using DINO embeddings, XGBoost was trained for downstream classification on embeddings from the 65–90 hour time period. Model training and evaluation were performed using five times repeated 5-fold cross-validation, with class weights adjusted based on inverse class frequencies in scikit-learn (version 0.24.2). All other parameters followed the default settings in scikit-learn.

Attention Maps

To generate attention heatmaps, Gradient-weighted Class Activation Mapping (Grad-CAM) [128] was applied. Heatmaps were generated for StembryoNet at the time point with the highest probability of the embryo being classified as normal. The last feature extraction layer was used as the model's target layer, and the ETiX-embryo ground truth label was backpropagated.

Random Classifier

To benchmark against the models, a random classifier was simulated by sampling values from a uniform distribution between -5 and 5, followed by a sigmoidal transformation to obtain predicted probabilities for each embryo in the test set. This random classifier was evaluated using twenty times repeated 5-fold cross-validation.

3.2.2. Early-Stage Classification

For AI-based early-stage classification, two types of models, ResNet18 and a Support Vector Machine (SVM), were trained using both brightfield and fluorescence images. Since the goal was to forecast whether an ETiX-embryo would develop normally by 90 hours, embryologist annotations at that time point were used as target labels (Figure 3.5a). To evaluate model accuracy and determine key classification features over time, models were trained at 5-hour intervals from 0 to 90 hours, generating 19 distinct model sets (Figure 3.5a). While the ResNet models ($\text{ResNet}_{\text{BF}}$, $\text{ResNet}_{\text{Fluor}}$) were trained directly on the image data, the SVMs (SVM_{BF}, $\text{SVM}_{\text{Fluor}}$) were trained on features extracted from these images (Figure 3.5b).

ResNet

$\text{ResNet}_{\text{BF}}$ and $\text{ResNet}_{\text{Fluor}}$ were trained with the same number of epochs, batch size, optimizer, loss, and image augmentation as the StembryoNet backbone, but on unsynchronized data.

SVM

SVM_{BF} and $\text{SVM}_{\text{Fluor}}$ used a linear kernel for simplicity and interpretability of feature weights. Using a polynomial kernel resulted in overfitting, with the model predicting only the majority class, thus offering no performance improvement. As with ResNet, the SVM was trained and evaluated using five times repeated 5-fold cross-validation, stratified by ETiX-embryo class, with class weights adjusted based on inverse class frequencies. Z-score normalization was applied to the input data. To derive predictions, the SVM values from SVC.decision_function() were thresholded using a sigmoidal activation function, yielding better results compared to SVC.predict_proba().

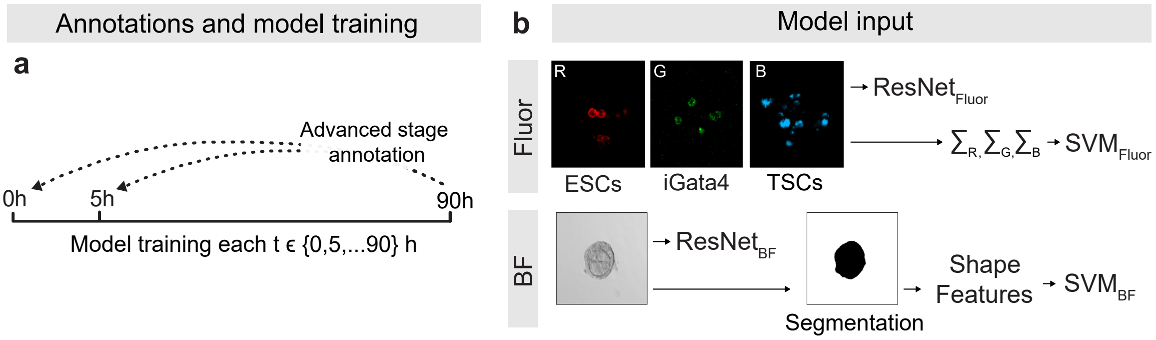


Figure 3.5: Annotations, model training, and model input for prediction of future normal ETiX-embryo development. (a) Diagram illustrating the approach to predicting future developmental outcomes of ETiX-embryos and identifying key classification features over time. Embryologist annotations from advanced-stage classification at 90 hours served as target labels. Models were trained at 5-hour intervals from 0 to 90 hours, using 5×5-fold CV for each interval. (b) ResNet and Support Vector Machine (SVM) models were trained on various types of images and features. ResNet models included $\text{ResNet}_{\text{BF}}$ and $\text{ResNet}_{\text{Fluor}}$, trained on brightfield and fluorescence images, respectively. SVM models were based on inferred ETiX-embryo characteristics, namely fluorescent intensity for $\text{SVM}_{\text{Fluor}}$ and shape features extracted from brightfield ETiX-embryo segmentations for SVM_{BF}. Adapted from Fig. 3 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout.

For interpretability of feature weights, a $\text{SVM}_{\text{BF+Fluor}}$ model was trained on brightfield and fluorescence features. Feature importance was determined using $\text{SVM}_{\text{BF+Fluor}}$ coefficients (referred to as *weights*), with absolute coefficient values used to quantify feature influence. SVM training and evaluation were conducted using scikit-learn (version 0.24.2).

The input to SVM_{BF} consisted of ETiX-embryo shape features extracted from brightfield segmentations (Figure 3.5b). Brightfield segmentation was performed using a SegFormer [111] model, trained on 450 selected brightfield images, which were sampled every 20 time points from 60 randomly selected ETiX-embryos and manually annotated. The training was conducted using the mmsegmentation GitHub repository [184] (version 0.30.0) with the following parameters: MiT-B0 architecture, pre-trained on ADE20k, a combination of Dice Loss and Cross Entropy Loss (weighted 10:1), AdamW (learning rate $= 1 \times 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay $= 1 \times 10^{-1}$), 1,000 training iterations. On-the-fly image augmentations included two steps: random flip (p=0.5) and z-score normalization (Equation 2.10). An 80%-20% split was used for training and testing, ensuring that images from the same ETiX-embryo were not present in both sets to prevent information leakage. The trained model was then used to infer segmentations for the complete dataset. Binary holes in the resulting segmentations were filled using the binary_fill_holes function from scikit-learn (version 0.24.2). In 0.2% of cases, no embryo was detected, primarily at early time points when individual cells had not yet formed compacted tissue. For these cases, missing values were imputed using the mean feature value at the corresponding time point. To quantify ETiX-embryo morphology and shape, PyRadiomics [114] (version 3.1.0) was used to extract the following 2D features: Elongation, Major Axis Length, Maximum Diameter, Mesh Surface, Minor Axis Length, Perimeter, Perimeter Surface Ratio, Pixel Surface, and Sphericity.

For $\mathrm{SVM}_{\mathrm{Fluor}}$, three fluorescence intensity values were used as input, corresponding to the sum of fluorescence intensities in the fluorescence in-focus image for each channel (Figure 3.5b).

Cell Counting

To count ESC, ESC-iGata4, and TSC cells for each ETiX-embryo at the cell-seeding time point, the cyto3 model from the Python package CellPose [109] was applied, using a cell pixel diameter of 7 on the corresponding fluorescence channel of the in-focus images. Each channel, representing a specific cell type, was processed separately for prediction. Following automated detection, the cell counts were manually reviewed and corrected by an embryologist.

3.2.3. Model Evaluation

Model performance was assessed using Accuracy and F1-score (F1). Training and evaluation were conducted using five times repeated 5-fold cross-validation, stratified by ETiX-embryo class (Figure 1.2). For each split, 75% of the training set was allocated for model training, while the remaining 25% was reserved for validation.

3.2.4. Morphological Trajectory

Morphological trajectory analysis was performed for both normal and abnormal ETiX-embryos. The analysis began with principal component analysis (PCA) on the nine PyRadiomics-calculated, z-score normalized (Equation 2.10) ETiX-embryo shape features, computed from 0 to 90 hours in steps of 5 hours, as previously described. A joint PCA was applied to project all embryos into a common low-dimensional space.

Morphological trajectory analysis was performed using a Python implementation [224] of the Slingshot algorithm [225]. Slingshot first constructs a minimum spanning tree among 19 clusters, each representing samples from distinct time points from 0 to 90 hours at 5-hour intervals, thereby establishing the global lineage structure. Using this structure, Slingshot fits a smooth principal curve through the PCA space, starting from the 0-hour cluster, which serves as the root. Pseudotime values are then computed by projecting each sample onto the fitted curve, yielding a continuous, one-dimensional measure of developmental progression along the trajectory. Two fitting epochs were used to ensure convergence of the principal curves. Trajectories were inferred separately for normal and abnormal ETiX-embryos to enable class-specific interpretation of developmental dynamics.

3.2.5. Improving Cultivation Efficiency

To determine the number of predicted normal and abnormal ETiX-embryos in datasets with increased initial cell counts—$\mathrm{Data}_{2\mathrm{X}}$ (twofold) and $\mathrm{Data}_{3\mathrm{X}}$ (threefold)—predictions from 25 StembryoNet models trained using five times repeated 5-fold cross-validation were applied to ETiX-embryos from Data$_{2X}$ and Data$_{3X}$ at 35 to 64 hours. Given the accelerated development of ETiX-embryos in Data$_{2X}$ and Data$_{3X}$ (Figure 3.6), 64 hours was selected as the final time point for evaluation of these datasets.

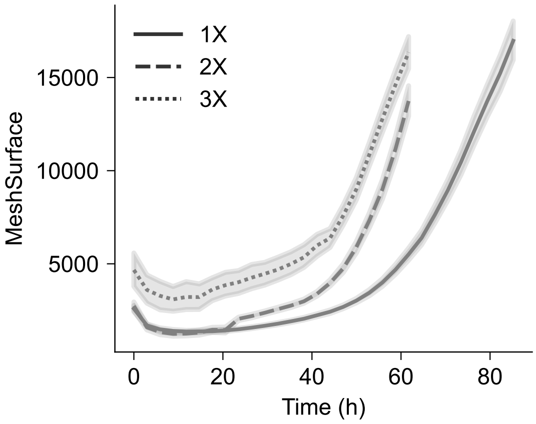


Figure 3.6: ETiX-embryo growth for Data$_{1X}$, Data$_{2X}$, and Data$_{3X}$. Higher initial cell numbers accelerate embryo growth. Error bars represent 95% confidence intervals across normal ETiX-embryos: 1X, $n = 206$; 2X, $n = 85$; 3X, $n = 173$. Adapted from Fig. S5 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout.

3.2.6. Analyzing Development of Individual ETiX-Embryos

To analyze ETiX-embryo development, model-predicted probabilities for the normal class were tracked for individual ETiX-embryos throughout the observation period. Therefore, a ResNet18 was trained using five times 5-fold cross-validation on images from every 5 hours. For quantification, ETiX-embryos were categorized into four groups: 1) continuously normal, 2) abnormal to normal, 3) continuously abnormal, and 4) normal to abnormal. Categorization was based on the predicted probabilities at 0 hours and 90 hours, combined with the slope of a fitted linear regression across all time points (Table 3.2).

Since model-predicted probabilities were observed to fluctuate for some ETiX-embryos, the slope of the linear regression was incorporated into the classification criteria to ensure more robust categorization. ETiX-embryos that did not meet any of the defined criteria in Table 3.2, due to ambiguous or more complex trajectories, were placed into an *Other* class to avoid making inaccurate statements about uncertain or more complex cases (Figure A.13). In total, 23.3% of the ETiX-embryos were assigned to the *Other* class, while the remaining embryos were assigned to the four defined categories. Qualitative evaluation of embryos assigned to the *Other* class suggested that these represented ambiguous cases that were also difficult to evaluate for an expert embryologist, especially given the high general variability observed in ETiX-embryo development. Further expanding the dataset may either help to reduce fluctuations in prediction trajectories and increase overall model certainty or uncover subtle, transient morphological patterns underlying these more complex dynamics.

Table 3.2: Criteria for ETiX-embryo categorization (Data$_{1X}$).

ETiX-embryo category	P(Normal) at 0 hours	P(Normal) at 90 hours	Regression slope
Continuously normal	≥ 0.5	≥ 0.5	≥ 0
Abnormal to normal	< 0.5	≥ 0.5	≥ 0
Continuously abnormal	< 0.5	< 0.5	< 0
Normal to abnormal	≥ 0.5	< 0.5	< 0

To categorize Data$_{2X}$ and Data$_{3X}$ (Figure 3.15d), the same ResNet models trained on Data$_{1X}$ were applied, using 5-hour intervals from 0 to 90 hours. Due to the accelerated development of ETiX-embryos in Data$_{2X}$ and Data$_{3X}$, prediction time points were adjusted to correspond to equivalent embryo sizes, ensuring comparability (Table 3.3, Figure 3.6).

Table 3.3: Time points used for Data$_{2X}$ and Data$_{3X}$ categorization. Time points for Data$_{2X}$ and Data$_{3X}$ were aligned based on corresponding developmental stages determined by an embryologist. The first four time points are identical, as phenotypes are comparable, for example, at 0 hours, all embryos consist of single cells undergoing aggregation, a process that continues over the initial time points. Later, development accelerates, as also reflected in ETiX-embryo growth (Figure 3.6). The later time points were selected by an expert embryologist, guided by both similar embryo size and developmental stage.

	Hour																	
Training (Data$_{1X}$)	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85
Prediction (Data$_{2X}$/Data$_{3X}$)	0	5	10	15	15	20	25	29	35	41	42	44	50	52	55	57	59	61

3.3. Results

3.3.1. Advanced-Stage Classification

Comparative Analysis

StembryoNet was compared to the state-of-the-art deep learning models ResNet18, ResNeXt, MobileNet, GoogleNet, and DenseNet, all trained on ETiX-embryo images captured at 90 hours, as well as a Multiscale Vision Transformer (MViT) trained on videos of ETiX-embryo development from 65 to 90 hours. To ensure an unbiased performance estimate, five times repeated 5-fold cross-validation was applied (Figure 1.2). Detailed information on model training and evaluation is provided in Section 3.2.

While a baseline random classifier achieved an accuracy of 50% (F1-score = 31%), StembryoNet, ResNet, and MViT all exceeded this baseline (Figure 3.7a). Notably, StembryoNet achieved a mean accuracy of 88% (F1 = 77%), outperforming ResNet (80% accuracy, F1 = 67%) and MViT (81% accuracy, F1 = 68%). The confusion matrix for StembryoNet indicates a recall of 83% for normal ETiX-embryos at a precision of 71% (Figure 3.7b). In comparison, ResNet yielded more false positives, resulting in a recall of 87% and a precision of 55%

for normal ETiX-embryos (Figure 3.7b). StembryoNet consistently outperformed not only ResNet and MViT but also other advanced deep learning models, demonstrating significant superiority across comparisons (Figure 3.7a). All models were trained on fluorescence in-focus images, as adding the brightfield channel did not enhance performance, and fluorescence in-focus images showed a superior outcome compared to fluorescence z-sum projection images (Figure 3.7c).

Testing StembryoNet on synchronized data (pre-selected time points) did not improve performance, indicating that StembryoNet eliminates the need for human intervention in selecting time points (Figure 3.7d). To determine whether training on synchronized data explains StembryoNet's superior performance compared to ResNet, its accuracy was evaluated on ETiX-embryos at 90 hours, achieving 79% accuracy (Figure 3.7d). These results suggest that the performance difference is not solely attributed to training on synchronized data but rather to StembryoNet's ability to fuse model predictions on unsynchronized data.

Fully supervised methods like StembryoNet rely on labor-intensive human annotations, which makes self-supervised, annotation-free deep learning approaches an appealing alternative. DINO, a recent self-supervised deep learning method, trains vision transformers to learn meaningful visual representations without labeled data. Although DINO performed inferiorly to StembryoNet, it was able to cluster normal and abnormal ETiX-embryos to some extent, showcasing the potential of annotation-free approaches for the future (Figure A.10).

The StembryoNet approach employs a straightforward classification method by selecting the maximum prediction from a time course and applying simple thresholding. In contrast, long short-term memory (LSTM) networks provide a more advanced deep learning strategy for time-series classification. To leverage temporal information, the StembryoNet backbone, which was trained on synchronized data, was used to generate raw predictions across the final 25 hours of ETiX-embryo development, which then served as input for the LSTM. The results showed that the LSTM significantly outperformed the ResNet, but StembryoNet significantly outperformed the LSTM (Figure A.11). This highlights the feasibility and effectiveness of the StembryoNet approach.

StembryoNet Cross-Testing

Model cross-testing provides valuable insights into model robustness and potential data biases. Training and testing StembryoNet on the entire dataset using repeated cross-validation resulted in the second-highest accuracy and the highest F1-score compared to using one dataset as a holdout test set (Table 3.4). This outcome is expected, as larger training sets typically improve performance. However, test performance varied across the three holdout datasets, with the model tested on dataset B exhibiting the most pronounced drop, a 0.09 decrease in F1-score compared to the best holdout test performance. This suggests underlying differences between the datasets, potentially due to variations in stem cell model cultivation or imaging conditions, which could impact feature extraction and classification accuracy.

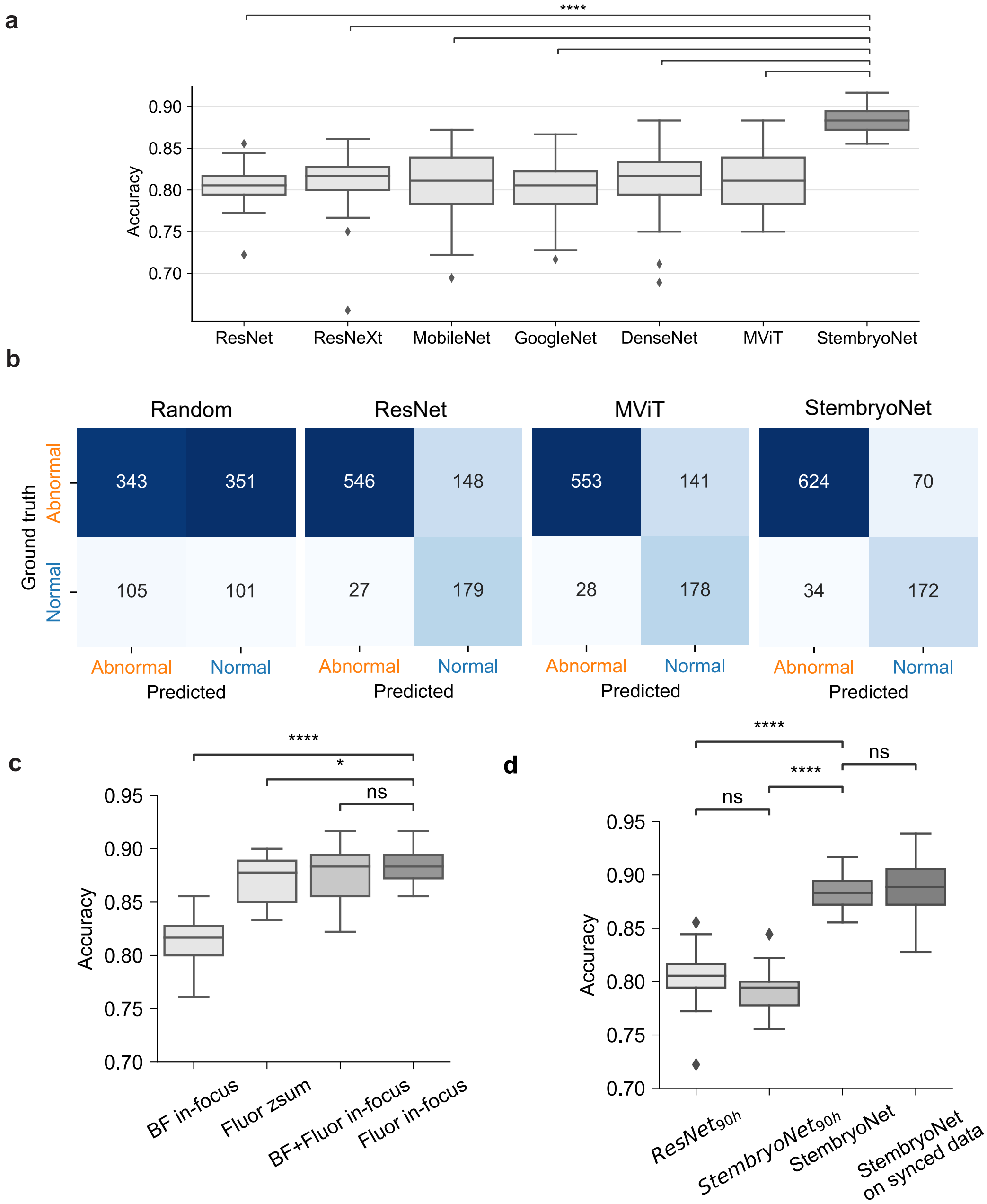


Figure 3.7: AI-based ETiX-embryo classification at advanced stages of development. (a) Performance comparison of StembryoNet to different classifiers that are trained on single-time points (ResNet, MobileNet, ResNeXt, GoogleNet, DenseNet) and multiple-time points: MViT, over five times repeated 5-fold cross validation (CV). (b) Confusion matrices of random classifier, ResNet, MViT, and StembryoNet averaged across 5×5-fold CV. (c) Comparison of StembryoNet classification performance using brightfield in-focus images, fluorescence z-sum projected images, combined brightfield and fluorescence in-focus images, and only fluorescence in-focus images. (d) Classification performance for 1) ResNet and 2) StembryoNet as described in the main text, 3) StembryoNet$_{90h}$, trained on synchronized data and tested on images from 90-hours, and 4) StembryoNet tested on synchronized data, i.e., at the expert-selected time point. *p<0.05, ****p<0.0001, ns: not significant, two-sided Welch's t-test. Adapted from Fig. 2 and Fig. S3 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified content and layout.

Table 3.4: StembryoNet cross-testing. Datasets A, B, and C represent the three recorded datasets used in this study. The performance of StembryoNet on A ∪ B ∪ C is based on five times repeated 5-fold cross-validation, as reported previously. The remaining results are obtained by training StembryoNet on two datasets while using the third as a holdout test set, with performance variability assessed through five random 75% training – 25% validation splits within the training data. Accuracy and F1-score are reported as mean ± SD.

Data_Train	Data_Test	Accuracy	F1-score
A ∪ B ∪ C	A ∪ B ∪ C	0.88 ± 0.02	0.77 ± 0.03
B ∪ C	A	0.84 ± 0.04	0.76 ± 0.04
A ∪ C	B	0.84 ± 0.02	0.67 ± 0.08
A ∪ B	C	0.89 ± 0.03	0.72 ± 0.03

StembryoNet Explainability

To biologically contextualize StembryoNet's predictions, Grad-CAM heatmaps were used to visualize the model's focus areas on both normal and abnormal ETiX-embryos (Figure 3.8a,b). For normal ETiX-embryos, the model predominantly focused on the regions where the ESC-iGata4-derived tissue encircles the ESC-derived tissue (Figure 3.8a). This is also the region considered by embryologists to assess successful development. Conversely, for abnormal ETiX-embryos, the model's attention varied, focusing on disparate image parts (Figure 3.8b). This is comprehensible as abnormal development can manifest in various ways, such as a lack of cavitation, incorrect compartment formation, mispositioned compartments, or a combination of these factors.

In comparison to the LSTM-based approach, StembryoNet offers greater explainability through interpretation of the time points of maximum probability for the normal class. These time points showed strong correlations with the embryologist's synchronized time points (Figure A.12). This alignment indicates that StembryoNet predictions closely correspond to expert annotations of similar developmental time points, enhancing confidence in the model's reliability and interpretability.

Comparison Across Multiple Annotators

To evaluate StembryoNet's performance in terms of both accuracy and speed, it was compared to three embryologists. The results show that, when using two of the three embryologists as the ground truth, StembryoNet slightly outperformed the least accurate embryologist (Figure 3.9a,b). For the third embryologist, StembryoNet's performance is slightly lower than that of the least accurate embryologist (Figure 3.9c). In terms of speed, StembryoNet is 18 times faster than the embryologists (Figure 3.9d). Overall, these findings suggest that StembryoNet achieves accuracy comparable to that of human experts but with a substantial speed advantage, highlighting its practical utility for high-throughput applications.

In conclusion, StembryoNet significantly enhances the ability to classify ETiX-embryos at advanced developmental stages, achieving superior accuracy and precision compared to

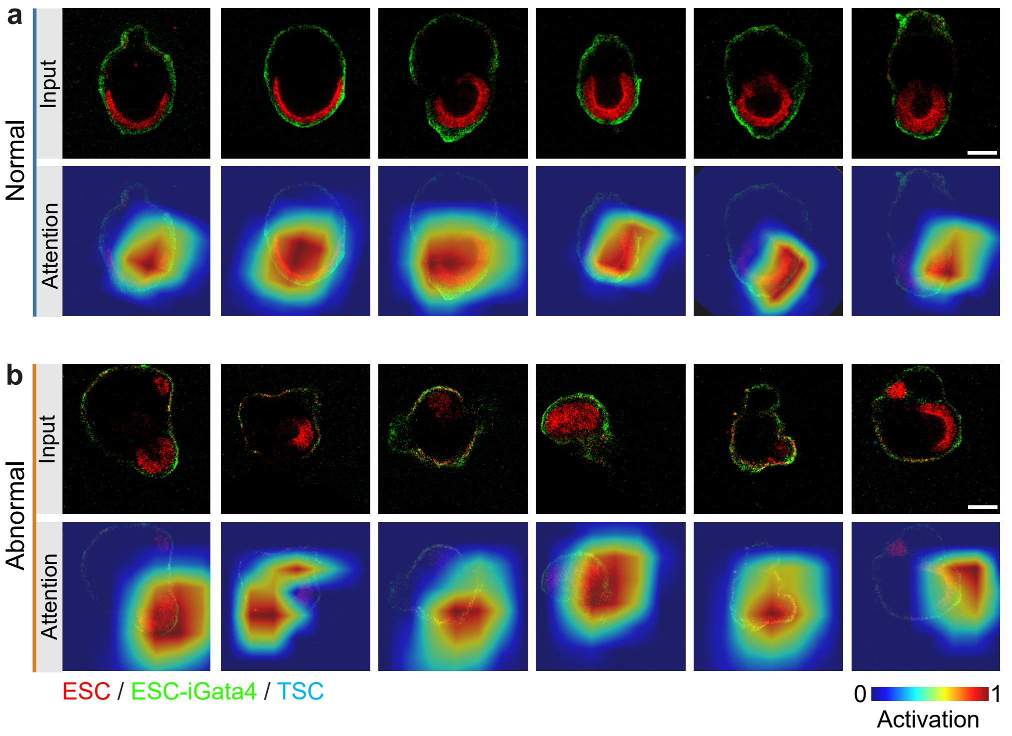


Figure 3.8: Grad-CAM attention maps highlight significant areas contributing to the classification decision of StembryoNet. (a) Normal ETiX-embryos (b) Abnormal ETiX-embryos. Scale bars: 100 μm. Adapted from Fig. 2 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout.

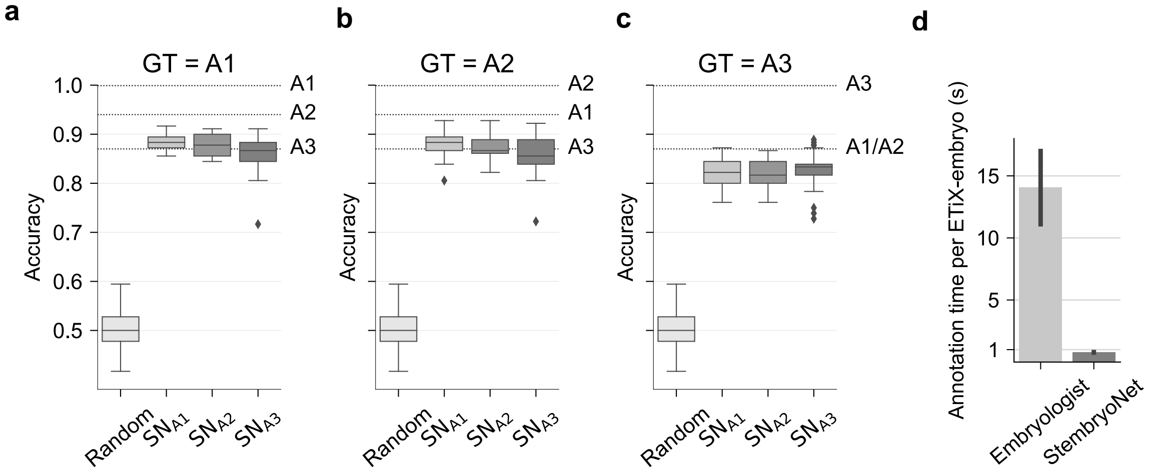


Figure 3.9: StembryoNet performance evaluation across multiple annotators. (a-c) Comparison of StembryoNet's performance with three embryologists (A1-A3). A1's annotations serve as the reference in the main text. GT represents the ground truth. $\mathrm{SN}_{\mathrm{A1}}$ – $\mathrm{SN}_{\mathrm{A3}}$ denote StembryoNet models trained on ground truth data from A1–A3. (d) Comparison of annotation time per ETiX-embryo in seconds between embryologists and StembryoNet. Error bars indicate 95% CI for embryologists ($n = 2$) and StembryoNet across 5×5-fold CV ($n = 25$). Adapted from Fig. S3 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout.

state-of-the-art deep learning models. By closely aligning with embryologist perceptions and providing biological insights, StembryoNet represents a robust tool for the analysis of ETiX-embryo development.

3.3.2. Early-Stage Classification

Deep-learning-based classification of normal and abnormal ETiX-embryos at 90 hours post-cell-seeding can improve the reproducibility of ETiX-embryo selection across laboratories. To advance this approach and gain deeper insights into the self-organization of normal ETiX-embryogenesis, two key objectives were pursued: (i) predicting ETiX-embryo development outcomes at earlier stages and (ii) identifying the features most predictive of successful development.

StembryoNet was not used for this task, as it is specifically designed for classification at later developmental stages, where ETiX-embryo synchronization, aligning time points based on comparable developmental stages, is feasible. A key strength of StembryoNet lies in its training on synchronized data. However, expert-based synchronization at earlier time points was not possible due to the lack of well-defined developmental criteria and the high variability observed during early development. To address this, three approaches were tested to transfer the synchronized time points from the advanced-stage classification task to the early-stage setting at 50 hours and train a StembryoNet model on it. The first approach involved a direct shift of the time window of the synchronized time points (65-90 h → 25-50 h, StembryoNet$_A$). The other two approaches additionally involved time-window shrinkage, linearly mapping 65-90 h to 35-50 h (StembryoNet$_B$) or to 45-50 h (StembryoNet$_C$). These strategies assumed that embryos exhibit less temporal desynchronization during earlier stages. However, all StembryoNet models did not improve results compared to a baseline model trained solely on unsynchronized single-time-point data (Table A.3). Consequently, the final approach was to proceed with training on unsynchronized single-time-point data rather than applying StembryoNet.

Two types of models were applied to both brightfield and fluorescence images: ResNet18 and Support Vector Machine (SVM), resulting in four model variations: ResNet$_{BF}$, ResNet$_{Fluor}$, SVM$_{BF}$, and SVM$_{Fluor}$ (Figure 3.5b). While ResNet models were trained directly on images, SVM models were trained on pre-extracted ETiX-embryo image features. Compared to the deep-learning-based ResNet18, SVM models provide greater explainability, as their manually defined features can be analyzed for importance. The performance gap between ResNet18 and SVM highlights the advantages of deep-learning-derived features over manually extracted ones. To ensure unbiased performance estimates, five times repeated 5-fold cross-validation was applied.

At the cell-seeding stage (t0), both SVM$_{Fluor}$ and ResNet$_{Fluor}$ significantly outperformed a baseline random classifier (Figure 3.10a). ResNet$_{Fluor}$ achieved an accuracy of 65% (F1 = 43%), while SVM$_{Fluor}$ performed comparably by achieving an accuracy of 64% (F1 = 42%). SVM$_{Fluor}$ feature weights indicated that the red (ESCs) and green (ESCs-iGata4) channels were most predictive (Figure 3.10b). In contrast, SVM$_{BF}$ performed significantly worse than random, and ResNet$_{BF}$ showed no improvement over random (Figure 3.10a), highlighting the critical role of fluorescence information — i.e., the fluorescent labeling of ESC, ESC-iGata4, and TSC cells — in early-stage predictions.

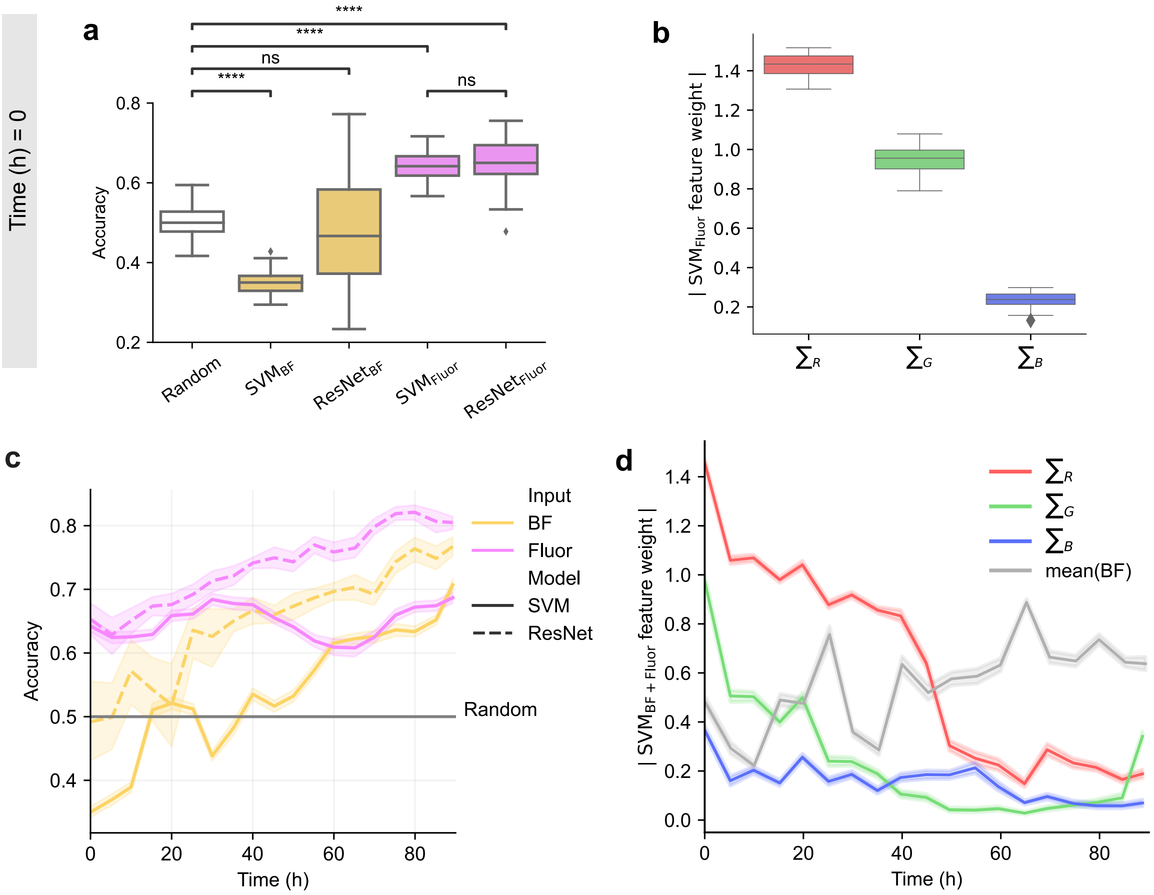


Figure 3.10: AI-based prediction of future normal ETiX-embryo development. (a) Classification performance of ResNet$_{\text{BF}}$, ResNet$_{\text{Fluor}}$, SVM$_{\text{BF}}$, and SVM$_{\text{Fluor}}$ at the time of seeding (t0). (b) Absolute feature weights of different channels of the SVM$_{\text{Fluor}}$ model at the time of seeding (t0). (c) Classification performance of all models along complete ETiX-embryo observation time, from 0 to 90 hours. (d) Absolute SVM feature weights of SVM$_{\text{BF+Fluor}}$ model throughout observation time. The feature importance of brightfield features was averaged. (c, d) Error bars represent 95% confidence intervals across 5×5-fold CV ($n = 25$). ****p<0.0001, ns: not significant, two-sided Welch's t-test. Adapted from Fig. 3 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout and colors.

As development progressed, classification accuracy improved compared to the cell-seeding stage (Figure 3.10c). ResNet$_{\text{Fluor}}$ consistently outperformed SVM$_{\text{BF}}$, SVM$_{\text{Fluor}}$, and ResNet$_{\text{BF}}$, particularly during later stages (i.e., 60 to 90 hours, Figure 3.10c). Over time, fluorescence sum values decreased in importance while ETiX-embryo shape features gained relevance (Figure 3.10d). In particular, ETiX-embryos with normal development followed a clear morphological trajectory (Figure 3.11b-e), displaying a higher perimeter and a lower PerimeterSurfaceRatio, indicative of a more compact shape, as opposed to the more fragmented appearance of abnormal ETiX-embryos likely due to tissue mispositioning (Figure 3.11a). Trajectory analysis was conducted using the Slingshot method [225], where ETiX-embryos from different time points (0 to 90 hours in 5-hour intervals) were clustered and ordered based on morphological similarity. The trajectory was inferred using nine ETiX-embryo shape features extracted from brightfield segmentations, starting from the initial time point (0 hours). The analysis also captured radial symmetry breaking around 65 hours, where normal ETiX-embryos, initially spherical with radial symmetry, began to elongate, while abnormal ones maintained similar sphericity (Figure 3.11a).

In summary, deep-learning approaches can predict ETiX-embryo success from the cell-seeding stage, with fluorescence data as the key predictor. As development progresses, prediction accuracy increases, and morphological features gain predictive importance.

3.3.3. Analyzing Development of Individual ETiX-Embryos

To assess individual ETiX-embryos and evaluate their developmental trajectories, distinct developmental patterns were identified: continuously normal, abnormal transitioning to normal, continuously abnormal, and initially normal-looking embryos that later exhibited abnormalities as development progressed (Figure 3.12a-d). For instance, one ETiX-embryo initially showed abnormal development and apparent lineage allocation failure but eventually readjusted after 35 hours post-cell-seeding, exhibiting a 20-hour delay compared to a normal embryo (Figure 3.12b,e). Another ETiX-embryo, which transitioned from normal to abnormal development, initially resembled a normal ETiX-embryo before deviating at around 35 hours post-cell-seeding, likely due to a failure in tissue sorting (Figure 3.12d,e).

This classification revealed a spectrum of developmental outcomes, ranked from most to least frequent: continuously abnormal, initially normal then becoming abnormal, continuously normal, and initially abnormal then turning normal (Figure 3.12f). This categorization provides deeper insight into ETiX-embryogenesis dynamics and highlights the necessity of tracking individual ETiX-embryo trajectories over time, rather than assuming that successful development is established early and maintained throughout the observation period.

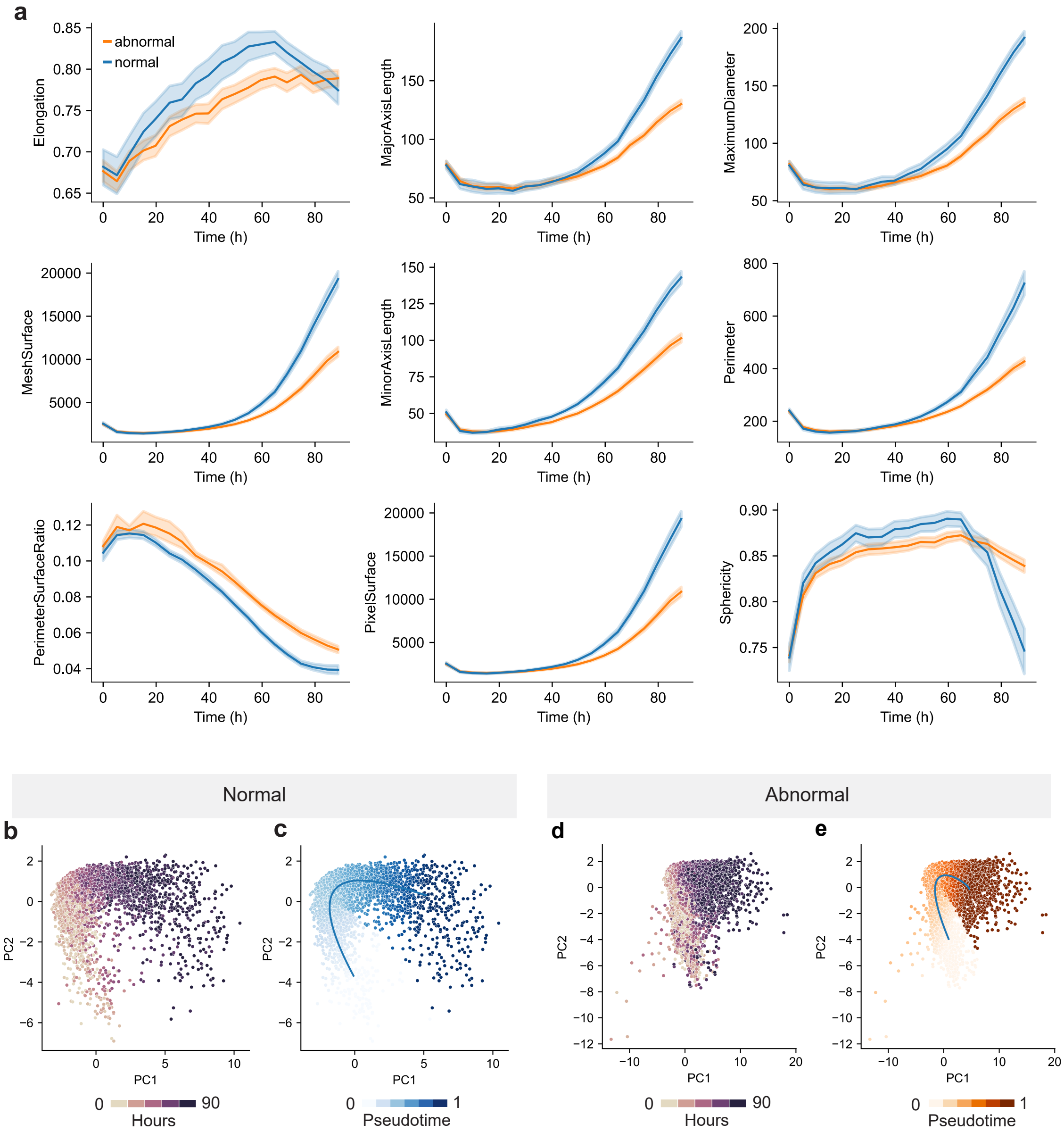


Figure 3.11: Morphology and morphogenetic trajectories of ETiX-embryos. (a) Nine ETiX-embryo shape features, based on brightfield ETiX-embryo segmentations, during development, from 0 to 90 hours in steps of 5 hours for all 900 ETiX-embryos. (b, d) Common Principal Component Analysis (PCA) of shape features for normal (b) and abnormal (d) embryos during the same time interval as in (a). The PCA was conducted on the combined dataset of normal and abnormal ETiX-embryos. The first two principal components (PC1 and PC2) explain 85.7% of the total variance. (c, e) Morphological trajectories of normal (c) and abnormal (e) embryos with associated pseudotime (Section 3.2.4). Error bars represent 95% confidence interval across abnormal ($n = 694$) and normal ($n = 206$) ETiX-embryos. Adapted from Fig. S6 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified colors.

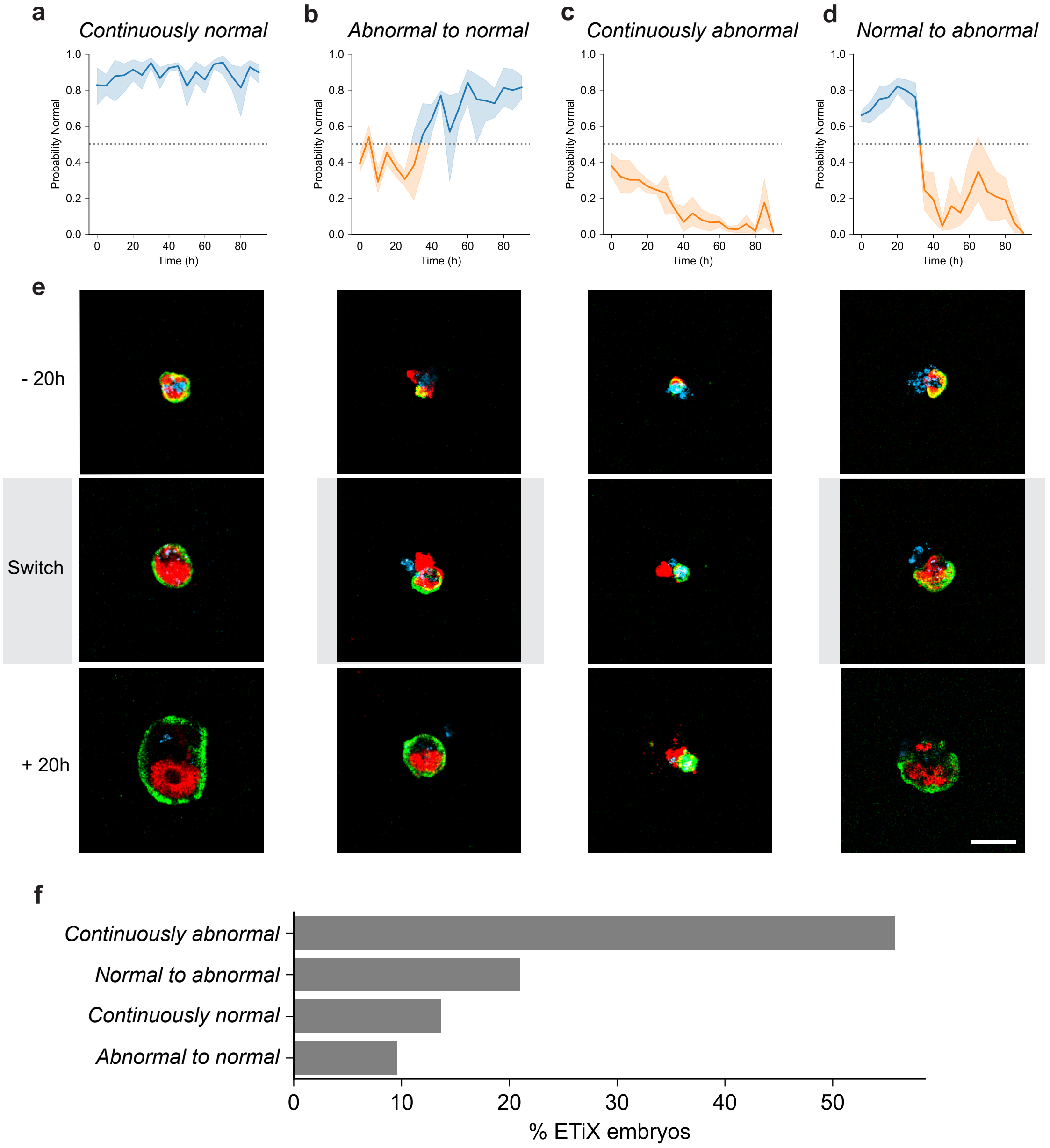


Figure 3.12: Time course ResNet predictions allow the identification of distinct ETiX-embryo developmental trajectories. (a, b) ResNet predictions of the probability of being classified as normal for two selected normal ETiX-embryos plotted over the observation period. For this application, ResNet$_{\text{Fluor}}$ was trained on unsynchronized data at distinct time points, from 0 to 90 hours in 5-hour intervals (Figure 3.5a,b), using five repeated 5-fold cross-validation runs (5×5-fold CV). (a) Continuously normal, (b) transitioning from abnormal to normal. (c, d) same as (a, b) but for two selected abnormal ETiX-embryos. (c) Continuously abnormal, (d) transitioning from normal to abnormal. (e) ETiX-embryos from (b, d) are shown at -20 hours and +20 hours relative to the model prediction switch, as well as at the moment of the switch. For (a), the same time point as in (b) is shown, and for (c), the same as in (d). (f) The proportion of embryos in different categories based on ResNet-predicted probabilities throughout the observation period. Scale bar: 100 μm. Error bars represent 95% confidence intervals across 5×5-fold CV in which the respective ETiX-embryo was part of the test set ($n = 5$). Adapted from Fig. 4 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified colors.

3.4. Pipeline Automation

The developed deep learning models for advanced- and early-stage classification of ETiX-embryos provide an automated approach to selecting embryos with normal development. However, for the main results, ETiX-embryo segmentation was performed manually to allow greater flexibility in selecting embryos, particularly at advanced stages where overlapping embryos posed a challenge. In this work, a workflow for automated ETiX-embryo segmentation, AI-based classification, and prediction integration into the Zeiss imaging software was implemented. This fully automated workflow is implemented and publicly available on GitHub at `https://github.com/deiluca/resnet_inference_zeiss`.

First, the brightfield mosaic image at the corresponding time point ($|X_{\text{mos}}| \times |Y_{\text{mos}}| \times |Z|$) undergoes a maximum intensity projection along the z-axis to generate a 2D image ($|X_{\text{mos}}| \times |Y_{\text{mos}}|$). Next, individual tiles from the 5×5 mosaic image are extracted (Figure 3.13a). Tile-level segmentation is then performed using the Segment Anything Model (SAM [112], Figure 3.13b). The tile-level segmentations are subsequently combined into a single image (Figure 3.13c). Tiles are merged based on their original grid positions, and touching objects at tile boundaries are consolidated into single objects (Figure 3.13d). To ensure accurate segmentation, objects at the outer edges of the imaging area are excluded as they might be incomplete (Figure 3.13d). Using the resulting mask, individual 5D ETiX-embryo samples are extracted by applying embryo-wise masking and five iterations of binary dilation to account for segmentation inaccuracies. A deep learning model is then employed to classify each ETiX-embryo as either normal or abnormal (Figure 3.13e).

Based on these segmentations and predictions, corresponding regions are identified and re-imported into the Zeiss imaging software for experimental use (Figure 3.13f). This enables seamless integration into experimental workflows and supports the early selection of ETiX-embryos with AI-assisted guidance.

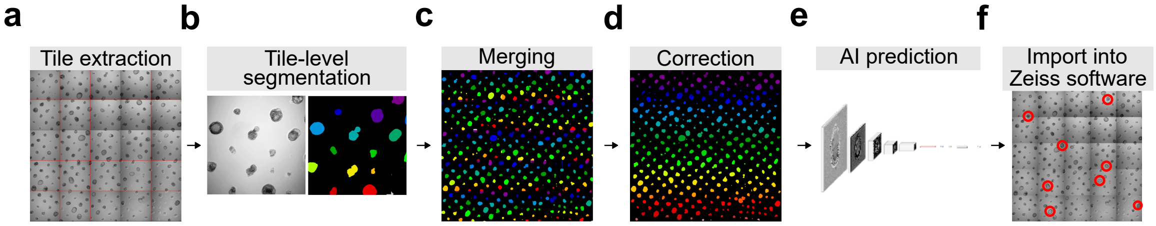


Figure 3.13: Pipeline for automated ETiX-embryo segmentation, AI prediction, and prediction import into Zeiss software. (a) Individual tiles are first extracted from the original 5×5 mosaic images. (b) Segmentation is then performed on each tile using the Segment Anything Model (SAM) [112]. (c) The segmented tiles are merged back into the original 5×5 mosaic grid. (d) Touching objects at tile borders are consolidated into single objects, and label enumeration is updated. (e) The StembryoNet model is then used to classify each ETiX-embryo as either normally or abnormally developed. (f) Finally, the locations of normally developed embryos are imported back into the Zeiss software, visualized as red circles on the original mosaic image. (b-d) Each detected ETiX-embryo is assigned a unique color.

3.5. Protocol Optimization

3.5.1. Overlapping ETiX-Embryos

Automated segmentation achieves high accuracy for intermediate to advanced stages of ETiX-embryo development (Figures 3.13). However, at very advanced stages, such as 90 hours post-cell-seeding, overlapping ETiX-embryos pose challenges for segmentation (Figure 3.14). One possible approach is to segment earlier time points when embryos are smaller and not yet in contact, followed by careful binary dilation of the masks. Alternatively, segmentations at very advanced stages should be reviewed and manually corrected.

For future studies, if very advanced stages are biologically relevant, it is advisable to prepare wells where ETiX-embryos are positioned slightly farther apart to facilitate automated segmentation.

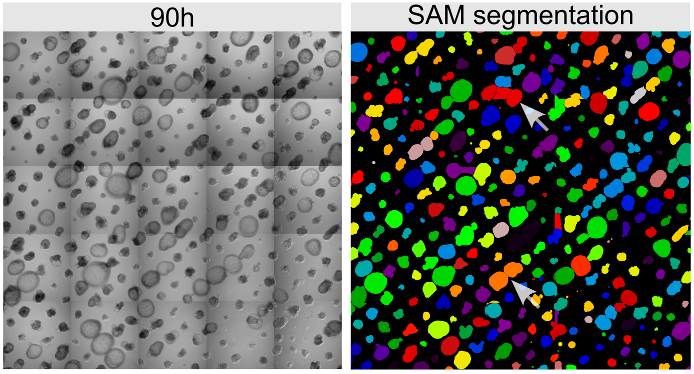


Figure 3.14: Automated ETiX-embryo segmentation using SAM becomes less accurate at later stages due to overlapping embryos. Brightfield image at 90 hours post-cell-seeding alongside the corresponding Segment Anything Model (SAM) segmentation. Each detected ETiX-embryo is assigned a unique color; touching objects with the same color indicates that they were mistakenly identified as a single embryo. Two such segmentation errors are highlighted with gray arrows.

3.5.2. Improving Cultivation Efficiency

Inherently, the channel-wise fluorescence sum (Figure 3.5b) is an estimate of the initial cell number and emerged as a distinguishing feature of ETiX-embryos with normal and abnormal future development. The protocol aims for each ETiX-embryo to contain, on average, five ESCs, five ESCs-iGata4, and sixteen TSCs at the initial cell-seeding stage. However, achieving these exact cell numbers during seeding is challenging due to the stochastic nature of cell distribution in each well, intrinsic to the experimental protocol. This variability could explain the differences in cell counts observed between normal and abnormal ETiX-embryos. Indeed, at the time of seeding (t0), an SVM model trained on the initial number of ESC, ESC-iGata4, and TSC cells ($SVM_{cellcount}$) performed comparably to

$\text{ResNet}_{\text{Fluor}}$ (Figure 3.15a), indicating that the initial number of cells serves as a predictor for successful development. Analysis of cell counts showed that, on average, abnormal embryos had significantly fewer ESC cells (normal: 6.7 ± 2.9, abnormal: 4.5 ± 2.5, mean ± SD) and ESC-iGata4 cells (normal: 6.0 ± 2.7, abnormal: 4.1 ± 2.3, mean ± SD, Figure 3.15b).

To investigate this further, additional datasets were generated by doubling ($\text{Data}_{2\text{X}}$) and tripling ($\text{Data}_{3\text{X}}$) the initial cell numbers for all three cell types. This approach was taken to explore whether increasing the overall cell count could mitigate the variability during seeding and ensure that each ETiX-embryo receives the necessary complement of cells for normal development. By uniformly increasing the numbers of all three cell types, the likelihood of each embryo acquiring a sufficient number of cells was expected to improve, thereby increasing the proportion of embryos developing normally. For this purpose, StembryoNet was utilized to predict the proportion of normal and abnormal ETiX-embryos for $\text{Data}_{2\text{X}}$ and $\text{Data}_{3\text{X}}$.

The results confirmed this hypothesis, showing a higher proportion of normal ETiX-embryos as the initial cell counts increased, with $\text{Data}_{1\text{X}}$, $\text{Data}_{2\text{X}}$, and $\text{Data}_{3\text{X}}$ exhibiting normal ETiX-embryos proportions of 23%, 32%, and 60%, respectively (Figure 3.15c). This suggests that increasing the initial number of cells across all lineages may enhance developmental outcomes by ensuring that each ETiX-embryo receives the minimum necessary amount of each cell type. Additionally, deep-learning-based brightfield embryo segmentation showed accelerated ETiX-embryo growth for $\text{Data}_{2\text{X}}$ and $\text{Data}_{3\text{X}}$, allowing them to reach similar developmental stages earlier than those in $\text{Data}_{1\text{X}}$ (Figure 3.6). This was further supported by a negative correlation between the synchronized time points and the number of ESC cells ($r = -0.33$, Figure 3.15e) and ESC-iGata4 cells ($r = -0.27$, Figure 3.15f) in $\text{Data}_{1\text{X}}$, underscoring the need for synchronization. Applying the categorization from Figure 3.12 to $\text{Data}_{2\text{X}}$ and $\text{Data}_{3\text{X}}$ revealed higher proportions of normal ETiX-embryos, further supporting the previous findings (Figure 3.15d).

3.6. Discussion

This study represents a significant advancement in the application of deep learning for classifying and analyzing the experimental variability of stem cell-derived embryo models, marking the first investigation of its kind in the field. StembryoNet, a novel deep-learning model specifically designed for ETiX-embryo classification at advanced developmental stages (90 hours post-seeding), achieved 88% accuracy (F1-score = 77%). This performance significantly surpasses that of other state-of-the-art models, such as ResNet18 and MViT ($p < .0001$), confirming the robustness and reliability of StembryoNet in distinguishing between normal and abnormal embryonic forms. Notably, the simpler 2D ResNet performed comparably to the more complex video classification model MViT. One possible explanation is that MViT may require larger datasets to effectively identify key developmental time points relevant for classifying an ETiX-embryo as normal or abnormal. Given the moderate size of the dataset, the increased complexity of MViT did not yield superior performance, suggesting that its advanced temporal resolution may be more beneficial

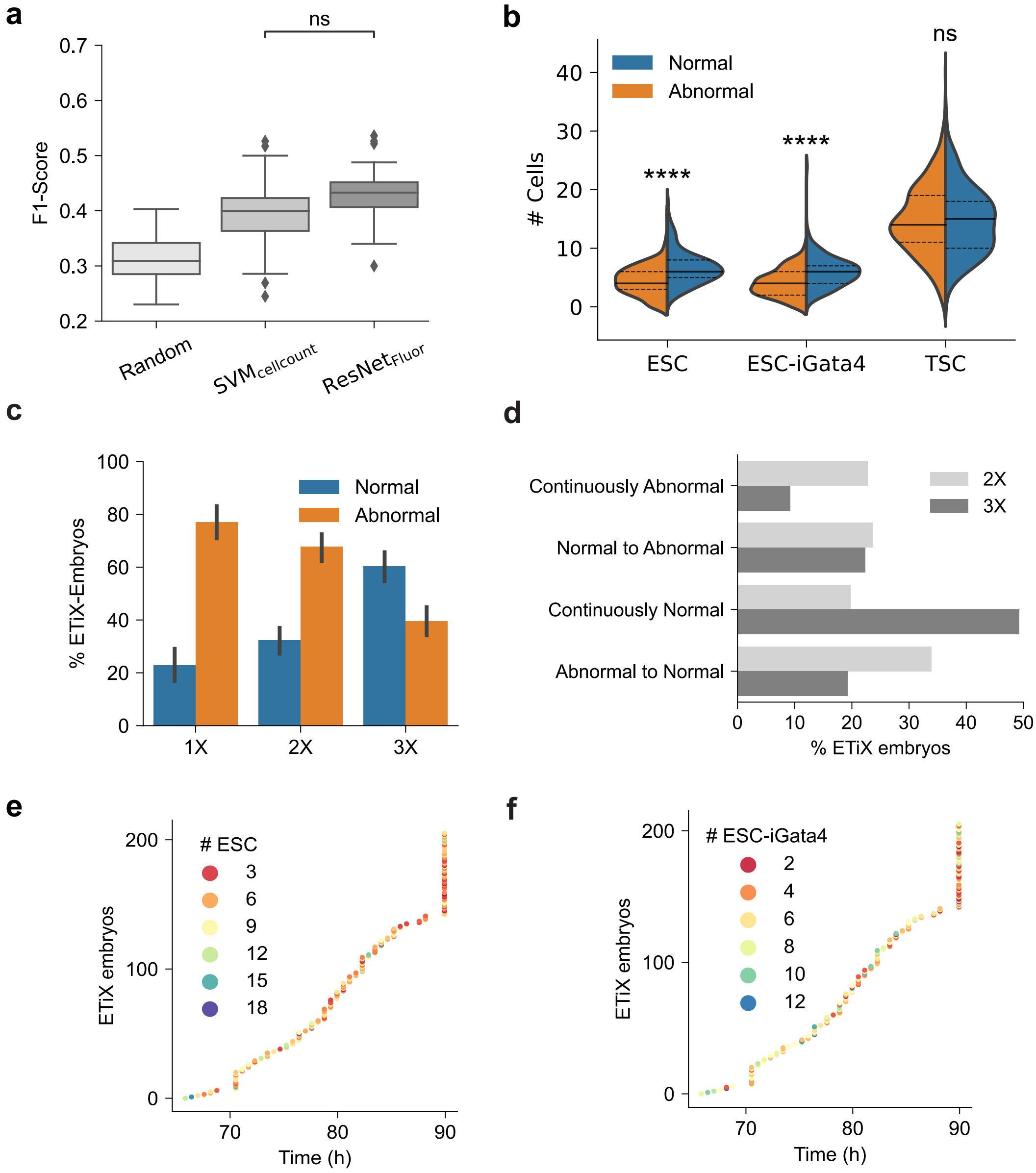


Figure 3.15: The initial ESC and ESC-iGata4 cell count is predictive of future normal ETiX-embryo development. (a) The number of cells at seeding is a strong predictor of subsequent normal development, consistent with predictions from ResNet$_{\text{Fluor}}$. (b) Normal ETiX-embryos have significantly more ESC and ESC-iGata4 cells at seeding, as shown by violin plots, where the horizontal lines indicate the 25th, 50th (median), and 75th percentiles. (c) Additional datasets with doubled (2X, $n = 306$) and tripled (3X, $n = 276$) initial cell counts of ESC, ESC-iGata4, and TSC cells showed a substantial increase in the proportion of embryos with normal development. Error bars represent 95% confidence intervals across three datasets (1X) and predictions from 5×5-fold cross-validation StembryoNet models ($n = 25$, 2X/3X). (d) Categorization of Data$_{2X}$ and Data$_{3X}$ ETiX-embryos into four classes: continuously abnormal, normal transitioning to abnormal, abnormal transitioning to normal, and continuously normal. (e-f) The number of ESC and ESC-iGata4 cells at seeding is negatively correlated with the expert-defined synchronization time point. $^{*}p<0.05$, $^{****}p<0.0001$, ns: not significant, two-sided Welch's t-test. Adapted from Fig. S5 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified colors.

for larger datasets. Currently, StembryoNet is limited to classification at advanced stages, as embryo synchronization – where an embryologist annotates the time point of similar development – is only feasible at that stage. Synchronization at earlier time points may enable StembryoNet to perform early classification in the future.

At the cell-seeding stage, the best-performing deep learning model achieved a classification accuracy of 65% (F1=43%), significantly outperforming a random classifier ($p < .0001$). The strong influence of fluorescence data at this stage suggests that early developmental predictions are highly dependent on initial cell count. Increasing the number of ESC, ESC-iGata4, and TSC cells by two- and threefold led to a higher proportion of normal ETiX-embryos, highlighting the importance of precise cell seeding techniques. However, larger datasets and enhanced imaging techniques—such as more frequent imaging intervals and higher spatial resolution—may reveal subtle developmental cues in brightfield images at the cell-seeding stage that correlate with successful outcomes.

As ETiX embryogenesis progresses, raw fluorescence data become less informative, and morphological characteristics gain predictive importance. Normal ETiX-embryos follow a distinct morphological trajectory, characterized by overall larger sizes, more compact shapes in early stages (i.e., from 15 to 60 hours), and less spherical shapes post-radial symmetry-breaking (i.e., from 65 hours onwards). During this period, $\text{ResNet}_{\text{Fluor}}$ outperforms $\text{ResNet}_{\text{BF}}$. During this period, $\text{ResNet}_{\text{Fluor}}$ consistently outperformed $\text{ResNet}_{\text{BF}}$, likely due to its ability to detect the formation of properly sorted tissue compartments, whereas $\text{ResNet}_{\text{BF}}$ remained limited to overall ETiX-embryo shape.

Previous work focused on manually selected ETiX-embryos around four to eight days after seeding for further molecular and cellular characterization, leaving the early stage of self-organization relatively unexplored. The AI models, trained at distinct time points from 0 to 90 hours, enable the selection of embryos at earlier stages when human selection is not feasible. For example, the model trained on data at 60 hours, achieving a classification accuracy of 76%, can select ETiX-embryos at the time of radial symmetry breaking for further analysis of this important developmental milestone. Furthermore, despite the predictive power of the cell count at seeding for normal development, ETiX-embryos with adequate initial cell setups can still fail at later stages, and vice versa. The deep learning model successfully identified such anomalies (Figure 3.12). Future studies could utilize single-cell RNA sequencing to explore the mechanisms underlying these developmental deviations, which are central to the system's highly self-organizing properties. Building on this concept, the deep-learning-based automated selection of well-developed ETiX-embryos could be integrated into microscope software, utilizing photoactivable dyes to select normal ETiX-embryos at any stage for further characterization.

The models developed here enable accurate classification of ETiX-embryos at both early and late stages of development. Incorporating insights from model predictions into experimental workflows, such as optimizing initial cell seeding and using StembryoNet to quantify protocol efficiency, has been shown to yield measurable improvements in cultivation outcomes. Building on this foundation, future high-throughput experiments could systematically investigate additional protocol parameters to further increase the consistency and quality of stem cell-derived embryo models.

To promote transparency, reproducibility, and community engagement, both the imaging datasets and all code required to train, evaluate, and test StembryoNet and related models have been made openly available on Zenodo (`https://zenodo.org/records/14605093`) and GitHub (`https://github.com/deiluca/StembryoNet`).

In summary, this study not only enhances the ability to reliably classify and predict ETiX-embryo development using deep learning but also deepens the understanding of the developmental dynamics involved. These insights pave the way for a reliable selection of ETiX-embryos for further research throughout the entire observation period, contributing to the emerging field of stem cell-derived models' embryology. Finally, the methodologies and findings presented can be extended to all stem cell-derived embryo models, both in mouse and human, broadening the impact of the work and opening new avenues for research and clinical innovations.

4. Implications for Measurement, Automated Analysis, and Protocol Optimization

In this work, stem cell model quantification was performed using multiple techniques, including magnetic resonance imaging (MRI), brightfield and epifluorescence microscopy, live imaging with confocal microscopy for brightfield and fluorescence capture, single-cell RNA sequencing (scRNA-Seq), and imaging flow cytometry (IFC). Each modality provides unique advantages and limitations, enabling comprehensive analysis tailored to different aspects of stem cell characterization. To analyze these datasets, automated pipelines were implemented to streamline data processing and support iterative protocol optimization.

Building on insights from Chapter 2 and 3, this chapter summarizes and discusses key implications for measuring stem cell-derived models, automating the analysis, and optimizing experimental protocols.

4.1. Measurement

4.1.1. Comparison of Data Modalities

Non-invasive measurement techniques allow continuous monitoring of stem cell-derived models without sample sacrifice, enabling time-series data collection. MRI, brightfield microscopy, and live imaging facilitate such non-invasive imaging (Table 4.1), whereas immunofluorescence imaging, scRNA-Seq, and IFC require sample destruction, limiting longitudinal studies. However, exposure to magnetic fields (MRI) or light (brightfield and fluorescence microscopy) may affect cell integrity and influence results [226, 227]. Therefore, while non-invasive techniques offer clear advantages, they also present limitations that must be carefully considered.

Spatial resolution is a key factor in imaging-based analyses. Compared to brightfield and fluorescence microscopy, MRI has 53-fold lower resolution (Table 4.1). For brain organoids, MRI achieves an edge length of 50 voxels, whereas for smaller ETiX-embryos, it would be reduced to only 3 voxels. Consequently, MRI is better suited for larger structures like brain organoids, while higher-resolution modalities remain essential for smaller stem cell models. Increasing MRI resolution requires stronger magnetic fields [228], while enhancing spatial resolution in live imaging reduces the number of samples that can be processed in parallel.

The time required per sample and the feasibility of parallel imaging affect both efficiency and temporal resolution. MRI acquisition is 78 times slower than brightfield and fluorescence microscopy and 111 times slower than live imaging (Table 4.1). In contrast, live imaging enables automated time-series recording, reducing manual effort compared to conventional microscopy. For non-automated methods, a trade-off exists between imaging frequency and analytical value, while exposure to magnetic waves (MRI) or light (microscopy) can impact sample integrity. In our studies, live imaging provided the highest throughput, capturing the largest number of parallel samples in the shortest time.

Sample visibility and optical distortion can influence measurement accuracy (Table 4.1). MRI ensures full sample visibility throughout imaging, whereas brightfield and fluorescence microscopy sometimes capture only partial sample views (Section 2.2.4.2). Live imaging maintained full ETiX-embryo visibility except at imaging borders or in overlapping samples toward the movie's end (Section 3.5.1). Brightfield imaging of brain organoids exhibited optical distortions due to well transfers and Matrigel embedding (Table 2.2.4.1), whereas live imaging, using a custom well throughout the observation period, eliminated such distortions. Overall, sample visibility and distortion are determined by the chosen protocol. From a measurement standpoint, using the same well for the entire observation period minimizes distortion, but samples must also be spaced sufficiently far apart to prevent overlap during later developmental stages.

Overall, MRI provides advantages in 3D volume capture but is limited by lower resolution and longer acquisition times. Live imaging demonstrated the shortest recording time per sample. Each data modality was tailored to the specific needs of brain organoids and ETiX-embryos, optimizing analysis for each study. Future research should aim to apply all modalities to the same biological samples to enable direct comparison. Recent protocols have also begun to employ live imaging for brain organoids [44].

4.1.2. Differences Between Stem Cell Models

The investigated stem cell models differ in cultivation duration, size, and measurement feasibility. Brain organoids require several months of growth, whereas ETiX-embryos develop within just four days [26, 229]. Continuous live imaging of brain organoids is impractical due to the prolonged use of high-tech equipment, which is typically shared among research groups. Size differences further impact measurement strategies: brain organoids can reach three to four millimeters in diameter [55], while ETiX-embryos measure only 200 μm by day 4 [40]. This disparity affects imaging techniques like MRI, where lower resolution becomes a limiting factor for smaller samples, reducing the ability to extract detailed structural information.

Table 4.1: Measurement and automated analysis features of different data modalities. MRI, magnetic resonance imaging; BFM, brightfield microscopy; EFM, epifluorescence microscopy; LI, live imaging using confocal microscopy to capture both brightfield and fluorescence; scRNA-Seq, single-cell RNA-Seq; IFC, imaging flow cytometry; h/d: hours/days including sample preparation; ^{1}, T2*-w sequence; ^{2}, 10x magnification; ^{3}, edge length per voxel/pixel; ^{4}, diameter, 4,000 μm [55]; ^{5}, major axis length, 200 μm; ^{6}, based on our studies, but dependent on stem cell cultivation and imaging protocol; ^{7}, for time-lapse recordings; ^{8}, one image per organoid; ^{9}, due to sample thickness and marker unspecificity; ^{10}, cell morphology; ^{11}, number of cells × number of genes; ^{12}, number of cells × XYC. ✓, fulfilled; ×, not fulfilled; ~, partially fulfilled; NA, not applicable.

	Feature	**MRI^{1}**	**BFM^{2}**	**EFM^{2}**	**LI^{2}**	**scRNA-Seq**	**IFC**
Measurement	Non-invasive	✓	✓	×	✓	×	×
	Resolution^{3}	80 μm	1.5 μm	1.5 μm	1.5 μm	NA	0.5 μm
	Resolution for brain organoid^{4}	50 vx	2,600 px	2,600 px	2,600 px	NA	NA
	Resolution for ETiX-embryo^{5}	3 vx	130 px	130 px	130 px	NA	NA
	Recording time per sample	13 min	10 s	10 s	7 s	h/d	min
	Parallel recordings^{6}	3	1	1	300	32	1
	Entire samples visible	✓	~^{6}	~^{6}	~^{6}	NA	NA
	No optical distortion^{8}	✓	×^{1}	NA	✓^{6}	NA	NA
Automated analysis	Easy sample separation	✓	✓^{6,8}	✓^{6,8}	~^{6}	✓	✓
	Quality control	✓	✓	✓	✓	✓	✓
	3D volume	✓	×	×	~	NA	NA
	2D morphology	✓	✓	✓	✓	NA	✓
	Substructure morphology	×	×	~^{6,9}	✓^{6}	NA	~^{10}
	Cell count	×	×	~^{6,9}	~^{6}	✓	✓
	Cell distribution	×	×	~^{6,9}	~	✓	✓
	Gene expression	×	×	×	×	✓	×
	Data dimensionality	XYZ	XY	XYC	XYZCT	^{11}	^{12}

4.2. Automated Analysis

4.2.1. Comparison of Data Modalities

Different data modalities provide distinct analytical advantages and challenges (Table 4.1). Before automated analysis, images must be preprocessed to isolate individual samples. For the specific MRI protocol, rule-based extraction of single samples from three-sample images is possible (Section 2.1.2), while brightfield and fluorescence microscopy inherently capture one organoid per image, eliminating the need for separation. For live imaging, sample separation was straightforward in the early stages but became challenging as overlapping increased with growth (Section 3.5.1). These differences underscore how cultivation and imaging protocols shape the feasibility of automated analysis.

Quality control is crucial for optimizing imaging protocols, with each modality offering unique contributions. MRI and brightfield microscopy facilitate cysticity classification in brain organoids, with MRI enabling local cyst segmentation (Section 2.1.4.5). Live imaging allows distinguishing well-developed from abnormally developed ETiX-embryos (Section 3.2.1), highlighting how different modalities address quality control aspects specific to each stem cell model.

A major advantage of MRI is its ability to generate 3D images, enabling volume and morphology analysis, including brain organoid and local cyst segmentation (Section 2.1). In contrast, 2D morphology can be assessed across all imaging modalities (Section 2.2.3.3, Figure 3.11). However, MRI's limited spatial resolution restricts the detailed analysis of biologically relevant substructures such as ventricles in brain organoids.

Beyond overall morphology, quantifying biologically relevant substructures is essential. Fluorescence microscopy excels in imaging substructures, though ventricle-like structure quantification in brain organoids remains challenging, and neuron segmentation requires specific stains. Additionally, cell count quantification is hindered by sample thickness, causing blurred nuclear boundaries. However, these issues were not encountered in live imaging of ETiX-embryos in early stages, allowing accurate cell count estimation across three cell types. scRNA-Seq and IFC further refine cellular characterization by identifying cells based on gene expression profiles and fluorescence-based features, respectively.

Higher data complexity, such as increased temporal and spatial resolution or additional channel information, expands analytical possibilities. For ETiX-embryo analysis, optimizing z-plane selection and identification of the most informative imaging channel enhances downstream data analysis (Section 3.1). Although complex datasets demand greater computational resources and efficient handling, especially across geographically distributed research groups, this flexibility ultimately enhances analysis quality and insights.

4.2.2. Comparison of Classical and Deep Learning Models

Deep learning models are widely recognized as state-of-the-art for image classification and segmentation. However, in certain cases—particularly when data is limited (e.g., *Compactness* for MRI, Section 2.1.4.1) or when features are relatively simple (e.g., $\text{SVM}_{\text{Fluor}}$

at 0 hours, Section 3.3.2)—classical methods relying on feature engineering can achieve comparable or even superior performance. As dataset size increases and feature complexity grows, deep learning approaches tend to surpass classical techniques, as demonstrated in brain organoid segmentation in brightfield images (Section 2.2) and ETiX-embryo classification (Section 3.3). These findings highlight the importance of selecting the appropriate method based on data availability and feature complexity, ensuring optimal model performance.

Model explainability is crucial for interpreting predictions and assessing decision-making processes. In deep learning, explainability often relies on feature map analysis (Figures 2.15, 3.8), which provides a more abstract representation of model understanding. In contrast, classical methods based on engineered features may offer greater transparency, allowing insight into feature contributions. The principal component analysis applied to morphological features can highlight clone diversity for brain organoids (Figure 2.17) and class differences for ETiX-embryos (Figure 3.11). Additionally, support vector machines using manually defined feature sets can generate interpretable rankings of feature importance, improving the understanding of underlying patterns (Section 3.3.2).

4.2.3. Model Customization, Sanity Checks, and Generalizability

Deep learning customization spans multiple levels, from model selection to architectural refinement. Many published deep learning models offer varying levels of complexity, and in many cases, adopting a publicly available model by cloning its code repository, preparing data, and using default hyperparameters such as learning rate and image augmentation provides a solid starting point (Section 2.1, 2.2). However, model architecture is a crucial factor, as non-standard designs can lead to suboptimal performance (e.g., MOrgAna in Section 2.2), while architectural modifications can substantially enhance accuracy, as demonstrated for StembryoNet (Section 3.3.1). Model optimization is an iterative process that involves establishing baseline performance, identifying strengths and weaknesses, adjusting training data and hyperparameters, refining the architecture when needed, and continuously improving through repeated experimentation.

Deep learning models may perform close to random depending on task difficulty, making it challenging to distinguish between an incorrect model setup and a lack of meaningful signal in the data. Sanity checks can help diagnose potential issues, such as overfitting a model on a single training sample, where it should achieve near-perfect performance, to confirm its learning capability. Another approach is training the same model configuration on a different dataset with a known, simpler task to verify model functionality, as demonstrated for nuclei detection (Section 2.3.1). Additionally, testing multiple model architectures on the same task can help determine whether poor performance stems from model limitations or the inherent complexity of the classification problem, as illustrated for early-stage prediction of future successful development of ETiX-embryos (Section 3.3.2).

Model generalizability can be assessed through cross-testing, for example, by training on one dataset and testing on a separate dataset, rather than relying solely on cross-validation (Tables 2.2, 2.5, 3.4). Ideally, these datasets should originate from different laboratories or follow distinct experimental protocols to robustly evaluate the model's ability to generalize across varying conditions.

4.2.4. Model Evaluation

Cross-validation is a widely used technique for model evaluation, providing robust performance estimates and reducing the risk of overfitting. Leave-one-out cross-validation is well-suited for smaller datasets, such as the MRI (Section 2.1) and epifluorescence microscopy datasets (Section 2.3.2), while 5-fold cross-validation is preferable for larger datasets, including brightfield microscopy (Section 2.2) and live imaging (Chapter 3). Additionally, cross-validation enables uncertainty estimation by generating ensemble models from multiple validation splits, as demonstrated for epifluorescence ventricle segmentation (Section 2.3.2). To ensure unbiased generalization and prevent information leakage, each biological sample should be assigned exclusively to either the training or test set, particularly when data is collected across different labs (Section 2.2) or time points (Section 2.1). Furthermore, stratification should be applied to control for biological biases and confounding factors, such as clone origin (Section 2.2) or sample class (Chapter 3).

4.2.5. Differences Between Stem Cell Models

Differences between stem cell models influence the requirements for automated analysis. While features such as size, growth, and overall morphology are generally transferable across stem cell-derived models as demonstrated in Figure 2.17 and Figure 3.11, variations in measurement, including well types and sample grouping, can impact *in silico* sample separation (Section 3.1).

Quality assessment varies between stem cell models due to differences in structural development and temporal dynamics. Brain organoid quality is primarily evaluated based on cyst formation, which increases over time and leads to a gradual decline in structural integrity. In contrast, ETiX-embryos are assessed at the end of their four-day cultivation, with quality typically improving as structural organization progresses. Furthermore, the random distribution of starting cells in ETiX-embryos introduces developmental variability, requiring the development of a custom deep learning model called StembryoNet, which was trained on manually curated, time-synchronized datasets for phenotype recognition (Section 3.3.1).

These differences underscore the need for tailored automated analysis pipelines that account for stem cell model-specific characteristics, ensuring precise assessments and biologically meaningful insights.

4.3. Protocol Optimization

Automated analysis can be used to iteratively improve stem cell model measurement, cultivation efficiency, and reduce stem cell model variability.

4.3.1. Improving Measurement

4.3.1.1. Brightfield Microscopy

Incomplete sample visibility hinders accurate growth monitoring of stem cell models. By combining automated segmentation with the developed metric Edge Ratio, stem cell models that extend beyond the imaging borders can be effectively detected and quantified (Section 2.2.4.2). A lower Edge Ratio per recorded dataset indicates higher protocol quality, providing a reliable metric for evaluating cultivation protocols.

Optical distortion further complicates growth monitoring for complex stem cell model cultivation protocols, particularly those involving well transfers and different embeddings. Automated segmentation, as detailed in Section 2.2, facilitates the assessment of optical distortion across different wells and sample embeddings. Conducting such analyses requires systematic measurements where only one factor is varied at a time, allowing for the evaluation of individual variables contributing to optical distortion.

4.3.1.2. Epifluorescence Microscopy

Epifluorescence microscopy of sliced brain organoids encountered several challenges, including staining unspecificity and nuclei blurring due to overly thick slices (Section 2.3). Addressing these issues is critical for reliable downstream analyses.

Nuclei segmentation is essential for quantifying cell counts and distributions. However, large regional differences in segmentation performance necessitate recording z-stacks to improve nuclei visibility and segmentation accuracy.

Ventricle segmentation in brain organoids is vital for assessing neuronal development. Inter-model uncertainty could be leveraged for iterative protocol optimization, such as identifying the most specific immunofluorescence markers for ventricle staining (Section 2.3.2). This iterative process enhances the precision and reliability of ventricle identification, contributing to more robust developmental analyses.

Registration-based volumetric reconstruction of brain organoids provides comprehensive insights into their three-dimensional organization. Automated analysis pipelines could enable the systematic evaluation of protocol quality through key metrics, including the frequency of global tissue folds, the absolute and relative areas of local tissue folds, the occurrence of out-of-focus images, and the frequency of mirroring events. These quantitative assessments deliver valuable information on protocol performance and consistency, supporting iterative refinements in both organoid cultivation and analysis workflows.

4.3.2. Improving Cultivation Efficiency

As previously discussed, the first requirement is that the measurement process yields data of sufficient quality, avoiding critical issues such as samples extending beyond the field of view, which impairs interpretability. Once this is ensured, readouts from automated analysis pipelines can guide the iterative refinement of cultivation protocols, aiming to increase the proportion of high-quality structures and reduce variability in stem cell-derived models.

Cysticity represents an undesired differentiation pathway during brain organoid cultivation, necessitating accurate and automated measurement for quality monitoring and protocol optimization. Automated analysis tools, detailed in Section 2.1, enable both global cysticity classification and localization of cystic tissue. Robust global classification reliably distinguishes cystic from non-cystic organoids, while local cyst segmentation provides detailed insights into the spatial distribution and extent of cysts. These pipelines facilitate iterative comparative analyses, guiding protocol refinements to reduce cyst formation, improve organoid quality, and enable more reliable disease modeling.

For ETiX-embryos, established cultivation protocols achieve efficiencies of 22% [26]. To improve this outcome, the deep learning models distinguished well-developed from poorly developed samples. Initial analyses at the time of cell seeding revealed that normal ETiX-embryos exhibited a higher prevalence of specific cell types compared to abnormal embryos. This insight informed subsequent experiments, which involved increasing the number of seeded cells. Quantitative evaluation using deep learning models confirmed that refined protocols improved cultivation efficiency, with up to 60% of embryos classified as normal (Section 3.5.2). The developed deep learning model not only provides an objective measure of protocol success but also establishes a framework for further optimization toward more efficient cultivation protocols.

4.3.3. Decreasing Variability

Batch-to-batch variability remains a significant challenge in stem cell-derived models [42, 56, 230–232]. ScRNA-Seq and IFC enable the assessment of protocol variability by correlating cell counts across technical replicates, where higher correlations indicate lower variability and improved protocol quality (Section 2.4.1.4). Specifically for brain organoids, recent studies have identified mesenchymal cells as markers of low-quality brain organoids [210]. Their detection through scRNA-Seq or IFC provides a reliable means to distinguish between high- and low-quality brain organoids.

4.3.4. Differences Between Stem Cell Models

Sample visibility and optical distortion primarily stem from imaging and cultivation protocols rather than the specific stem cell-derived model. These challenges are influenced by factors such as well type and whether samples are imaged individually or in groups, both of which affect segmentation and *in silico* sample separation.

Fluorescence microscopy, commonly used for nuclei detection across all stem cell-derived models, faces similar technical limitations regardless of stem cell type. Many of these challenges can be mitigated through broadly applicable protocol optimizations. However, nuclei visibility is influenced by stem cell size: in smaller models, such as 2D cultures, sample slicing may not be necessary, as nuclei do not overlap in the z-direction.

Cultivation efficiency improvements also vary by stem cell model, as different factors define quality. For instance, cysticity serves as a key quality metric for brain organoids, while the structural organization of tissue compartments is crucial for ETiX-embryos. This underscores the need for tailored analysis pipelines that first quantify model-specific quality metrics before iteratively refining cultivation protocols.

4.4. Summary

This chapter outlined the broader implications of automated pipelines for the analysis, measurement, and protocol optimization of stem cell-derived models. The choice of imaging modality, sample preparation, and cultivation setup can introduce substantial variation in measurement resolution, visibility, distortion, and feasibility of automated analysis. Automated analysis pipelines offer a scalable and objective approach to assess these factors and enable iterative refinement of experimental protocols. By quantifying model-specific features, such as cysticity in brain organoids or successful organization in ETiX-embryos, automated methods support protocol optimization, improve cultivation efficiency, and reduce variability across batches, making them essential tools for the robust and reproducible analysis of stem cell systems. At the same time, model-specific characteristics, including size, structural organization, quality markers, and developmental dynamics, required tailored adaptations of the automated analysis pipelines to ensure reliable and meaningful assessments.

5. Conclusion and Outlook

> Section 5.2 is partially based on: L. Deininger, P. Caldarelli, M. Zernicka-Goetz, et al. "From pixels to patterns: The AI revolution in stem cell-derived models". Submitted to Nature Methods. 2025. [1]. The article was conceptualized and written by L.D. and P.C., with L.D. primarily responsible for the text and P.C. focusing on the figures.

5.1. Conclusion

Stem cell-derived models offer a groundbreaking approach to studying tissue development, function, and disease, providing physiologically relevant alternatives to human and animal tissues. However, manual analysis remains prevalent, despite being inefficient and susceptible to observer bias, underscoring the need for automated analysis pipelines. While automated approaches exist for certain imaging modalities, such as brightfield microscopy, gaps remain, for instance, the lack of pipelines for magnetic resonance imaging and challenges related to robustness against imaging artifacts in existing methods. This thesis developed automated analysis pipelines specifically designed for organ-level and organism-level stem cell-derived models across multiple imaging modalities, addressing these limitations and advancing quantitative assessment in the field.

Chapter 2 presents the development of automated analysis pipelines for brain organoids, an organ-level stem cell-derived model, integrating a diverse range of data modalities, including magnetic resonance imaging, brightfield microscopy, epifluorescence microscopy, scRNA-Seq, and imaging flow cytometry. Building on this foundation, Chapter 3 adapts the automated analysis approach to ETiX-embryos, an organism-level stem cell-derived model. Finally, Chapter 4 provides broader insights and recommendations into the measurement of stem cell-derived models, the automation of data analysis, and protocol optimization, offering a comprehensive comparison of different imaging modalities and computational strategies.

Summarized, the main contributions of this thesis are:

1. A novel automated analysis pipeline for brain organoids in MRI, enabling the tracking of organoid morphology with a Dice score of 0.92 and reliable differentiation between cystic and non-cystic organoids for quality control, achieving a ROC AUC of 0.98 (Section 2.1). The pipeline was further benchmarked against brightfield imaging to highlight its advantages and limitations relative to a simpler imaging modality.

2. A robust and explainable automated analysis pipeline for brain organoids using brightfield microscopy data, achieving a Dice score of 0.96. It enables reliable monitoring of organoid size, morphology, and diversity across imaging laboratories (Section 2.2). The pipeline incorporates strategies and quantitative metrics to identify and mitigate common measurement issues, including optical distortion and sample extension beyond imaging boundaries, to support the iterative refinement of measurement protocols.

3. Implementation of an automated analysis pipeline for brain organoids using epifluorescence microscopy, enabling the quantification of nuclei and ventricle-like structures. Furthermore, a registration-based volumetric reconstruction method based on serial slices of brain organoids was developed, along with strategies for iterative measurement optimization (Section 2.3).

4. Exploration of semi-automated scRNA-Seq data analysis and automated analysis of imaging flow cytometry data, demonstrating their potential for cell-level brain organoid analysis and quantification of variability in stem cell-derived models, supporting iterative protocol optimization (Section 2.4).

5. A novel automated analysis pipeline for quality monitoring of the organism-level stem cell-based ETiX-embryos at both early and advanced developmental stages, achieving an accuracy of 0.88. The pipeline incorporates a domain-specific learning strategy to account for developmental differences and enables the identification of features associated with successful development. It further supports the categorization of developmental trajectories, and provides quantitative insights to guide the refinement of cultivation protocols (Chapter 3).

6. Delineate the broader implications of automated quantification, measurement, and protocol optimization for stem cell-derived models, highlighting recommendations and best practices (Chapter 4).

7. To enable further development and usability of the proposed methods and data, the code and data of all published studies has been made publicly available as open data and open-source code (Section 2.1, Section 2.2, Section 2.4.2, Chapter 3)

In stem cell biology, the conventional approach has been to selectively continue cultivation with only the most successful stem cell-derived models. However, recent studies [210, 233–236], along with the present work, reflect a growing shift toward systematic protocol optimization aimed at improving model yield and reproducibility. The automated analysis pipelines developed in this thesis directly address key challenges in the field by enabling objective quality control, increasing cultivation efficiency, systematically quantifying variability across stem cell-derived models, and providing a comparative perspective on the strengths and limitations of different data modalities.

5.2. Outlook

As stem cell-derived protocols mature and organoid cultivation advances, a shift from optimization to large-scale, high-throughput screening is anticipated, enabling systematic disease investigation, compound screenings, and accelerated drug discovery. Currently, foundation models, large-scale AI models trained on diverse datasets to enable broad generalization and adaptability across multiple tasks, are making breakthroughs in segmentation tasks [112]. In the long run, these models are expected to evolve into versatile and generalizable tools, transforming the analysis of stem cell-derived models and advancing biomedical imaging as a whole.

To support such AI-driven pipelines, open-source repositories of annotated stem cell-derived datasets will be essential. These repositories not only allow public access to training data but also enable benchmarking and reproducibility, helping to identify methodological blind spots and improve generalizability across labs and protocols.

Another key direction is the full integration of automated analysis pipelines into experimental workflows. By detecting specific developmental patterns or morphologies as they arise, these systems can be used for active experimental steering. Researchers could dynamically intervene, e.g., selecting or perturbing specific structures in real time, enabling causal linkage between early morphological features and downstream molecular states. Such closed-loop systems would provide a powerful framework for uncovering the mechanisms of stem cell self-organization. In particular, the integration of such tools may lower the expertise required to operate advanced analysis pipelines, broadening accessibility and accelerating adoption in experimental laboratories.

At the same time, generative AI is demonstrating impressive capabilities across language, code, and increasingly, scientific domains. In stem cell research, it is already being used for tasks such as generating synthetic training data [237]. In the future, generative models could enable the prediction of experimental outcomes over time, either at the individual sample level, to predict specific developmental trajectories, or at the population level, to simulate system-wide responses under varying conditions. These tools may eventually allow researchers to replace certain experimental steps with *in silico* simulations, reducing time, cost, and resource consumption. Moreover, integrating text-based generative AI with image-based AI systems holds strong potential for cross-modal knowledge discovery and experimental planning, enabling automated hypothesis generation, protocol refinement, and literature-informed interpretation of results [238].

Finally, the integration of multimodal datasets, such as live imaging with transcriptomics or proteomics, will allow automated analysis pipelines to infer deeper structure–function relationships. Predicting molecular signatures directly from morphology, for example, could yield non-invasive markers for developmental potential or disease state, vastly increasing the interpretability and utility of stem cell-derived systems.

Together, these advances indicate a broader transformation: from *post hoc* image analysis toward automated experimentation in stem cell biology. By combining flexible architectures, standardized datasets, and real-time and multimodal modeling, future systems may not only interpret developmental processes but actively shape them.

A. Appendix

A.1. Analysis Pipelines for Brain Organoids

Table A.1: 3D U-Net training and inference for organoid segmentation. The times were measured using one NVIDIA GeForce RTX 3090 (24 GB) graphics card. Derivative of Table S4 by Deininger et al. [178] licensed under CC BY 4.0, with modified layout.

Model	**Training time (s)**		**Inference time (s)**	
	Per iteration	Total	Per sample	Total
3D U-Net	1.11	2,220	1.97	88.6

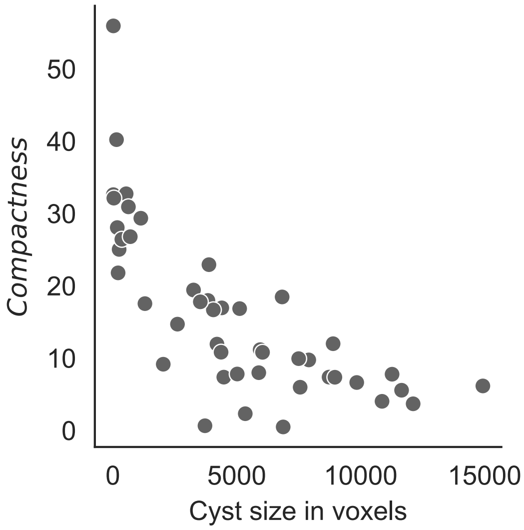


Figure A.1: *Compactness* negatively correlates with organoid cysticity. The cyst size in voxels is based on local cyst annotations.

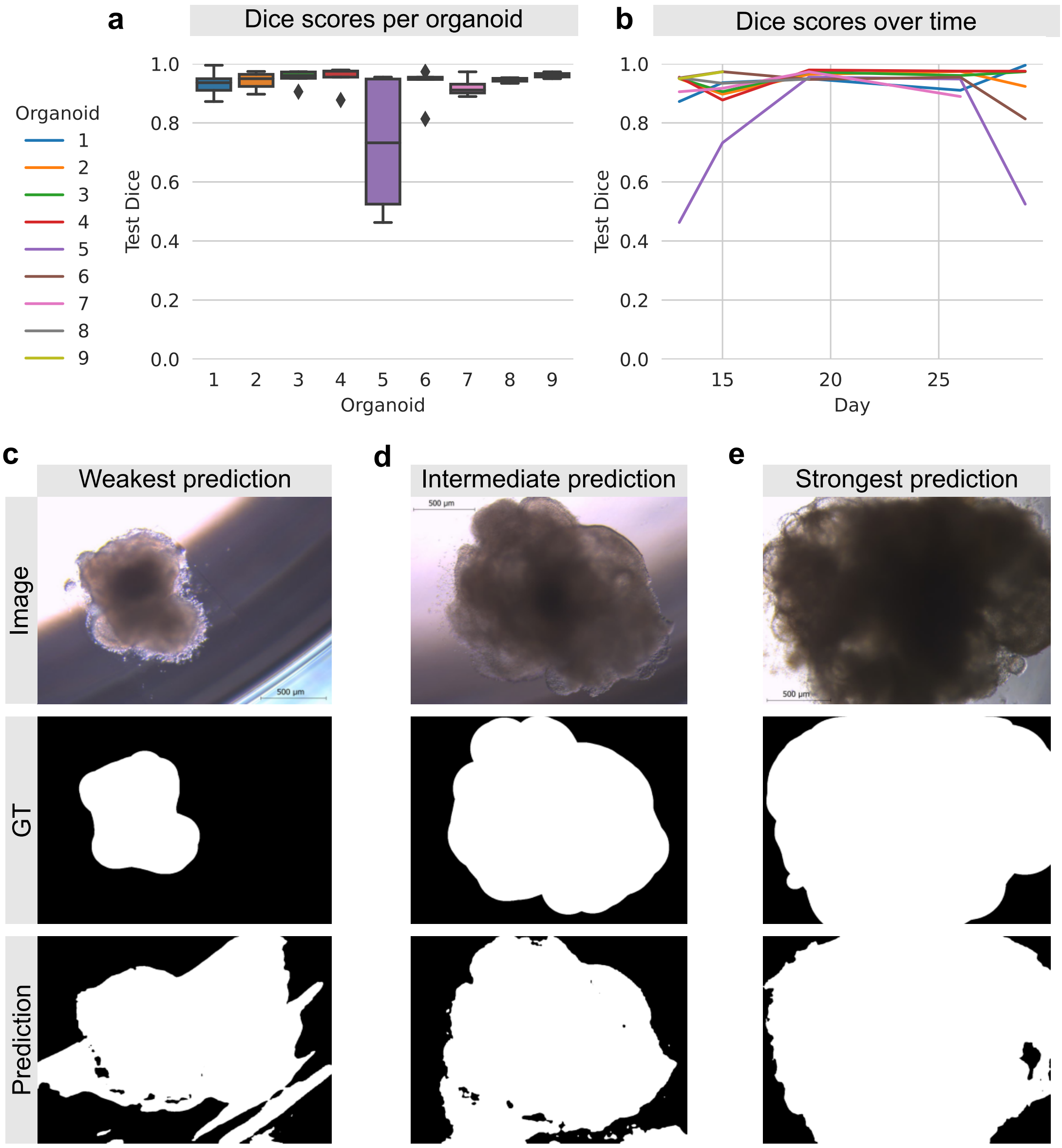


Figure A.2: 2D organoid segmentation for brightfield imaging. (a-b) Model performance. (c-e) Selected organoids based on model's prediction performance. (c) Organoid 5 (day 13): Dice score of 0.46. (d) Organoid 7 (day 15): Dice score of 0.92. (e) Organoid 4 (day 19): Dice score of 0.98. Image: original image, GT: Ground truth organoid location, Prediction: Predicted organoid location. Derivative of Fig. S1 by Deininger et al. [178] licensed under CC BY 4.0, with modified layout.

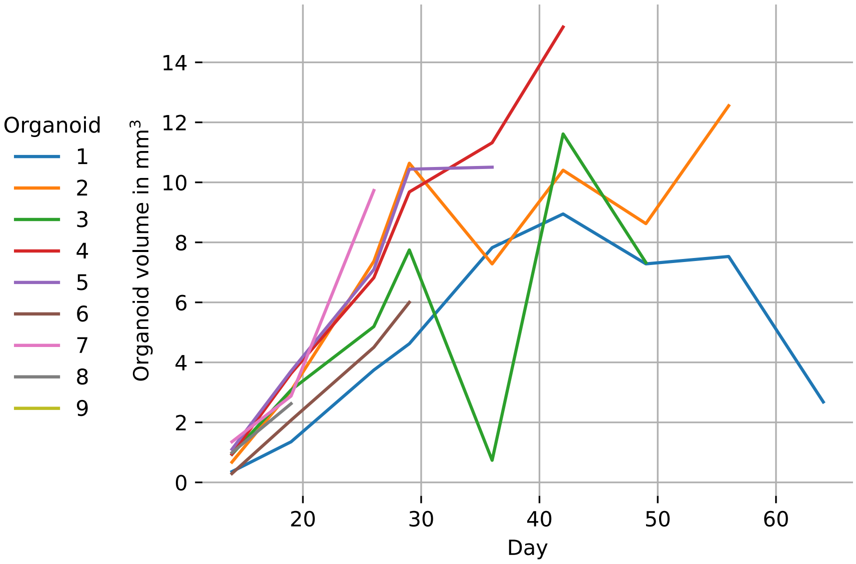


Figure A.3: Organoid growth over time. The organoid volume is based on the ground truth organoid annotation in the T2*-w sequence. Organoid 3 (day 36) has a sudden drop in volume which is due to the disruption of one or more cystic structures. Derivative of Fig. S2 by Deininger et al. [178] licensed under CC BY 4.0, with modified layout.

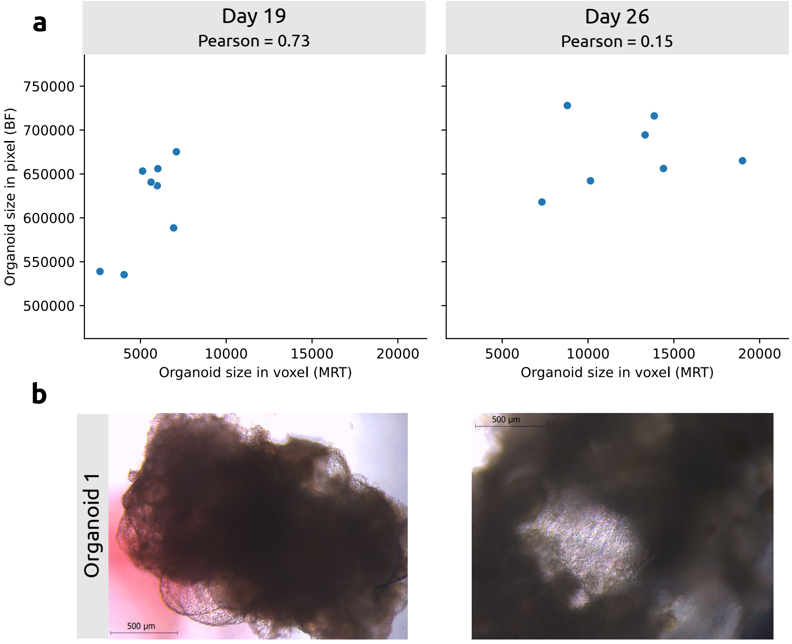


Figure A.4: Size correlation of the same organoids in MRI and brightfield microscopy. (a) Correlation of organoid sizes on day 19 and 26. For other time points, images were not acquired on the same day, complicating direct comparisons. Brightfield imaging was discontinued as organoids grew beyond the imaging boundaries. Organoid sizes based on ground truth organoid segmentations in the T2*-w sequence (MRI) and brightfield images. (b) Representative brightfield images of organoid 1 on days 19 and 26, illustrating that the organoid extended beyond the imaging field.

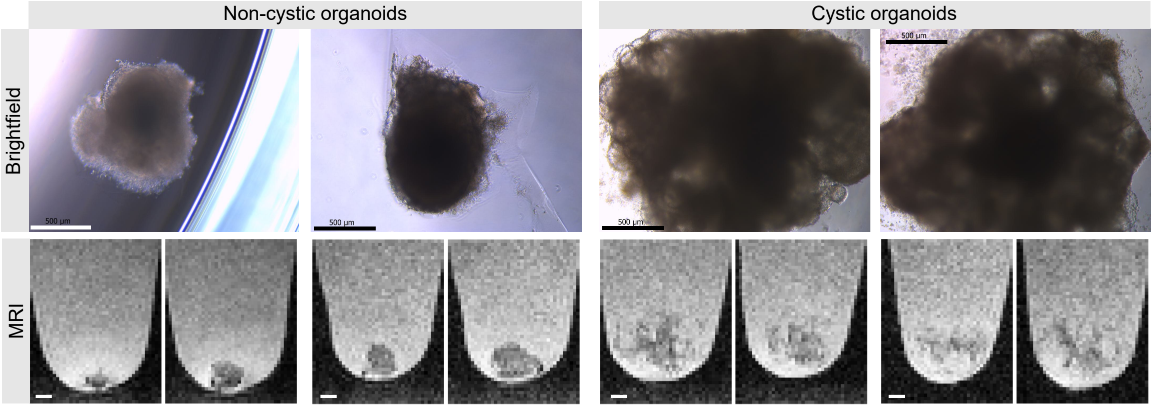


Figure A.5: Brightfield and MRI images of two non-cystic and two cystic organoids. The cystic organoids show fluid-filled cavities (or "cysts") and therefore resemble the same phenotype as reported in [56, 79]. Scale bars: 500 μm (brightfield), 400 μm (MRI). Images are from $Data_A$, organoid numbers (left to right): 1, 2, 4, 8. Derivative of Fig. S3 by Deininger et al. [178] licensed under CC BY 4.0, with modified layout.

Table A.2: Absolute and relative mean difference from ground truth organoid size. Best model for each imaging lab and day in bold. Max dev, maximal deviation. Adapted from Table S1 in Schroeter, Deininger et al. [198], licensed under CC BY 4.0, with modified layout.

		CellProfiler		OrganoSeg		$\text{MOrgAna}_{\text{MLP,W}}$		SegFormer	
	Day	abs (μm^2)	rel	abs (μm^2)	rel	abs (μm^2)	rel	abs (μm^2)	rel
Lab A	2	+1,297,875	+436%	+489,611	+163%	+250,687	+117%	-12,902	**-4%**
	5	+207,585	+21%	-22,707	-1%	+59,521	+7%	+222	**+0%**
	8	+404,867	+47%	+95,281	+11%	+188,018	+20%	+17,124	**+2%**
	10	+414,076	+44%	+136,576	+16%	+29,604	**+4%**	+60,645	+7%
	12	+294,134	+32%	+109,127	+10%	+182,146	+30%	+58,271	**+6%**
	16	+1,164,818	+133%	-36,974	-3%	+206,814	+18%	+13,877	**+1%**
	19	+685,425	+52%	-1,897	**0%**	+287,936	+21%	+17,595	+1%
	22	+648,911	+45%	-220,978	-13%	+445,152	+42%	+28,237	**+2%**
	25	+768,933	+41%	-433,490	-15%	+188,689	+7%	-26,257	**-2%**
	30	+456,924	+22%	-1,551,214	-44%	+205,316	+11%	-24,459	**-1%**
	Best		0/10		1/10		1/10		8/10
	Max dev.	+1,297,875	436%	+489,611	163%	+250,687	117%	60,645	7%
Lab B	2	+5,218,581	+1,768%	+839,366	+303%	+1,019,063	+351%	+9,667	**+3%**
	5	+1,229,865	+142%	+127,355	+14%	+327,160	+33%	-5,343	**-1%**
	8	+1,474,562	+173%	+238,197	+28%	+1,119,394	+137%	+41,078	**+5%**
	10	+2,989,813	+334%	+57,620	+8%	+1,454,328	+172%	+28,118	**+4%**
	12	+694,634	+72%	+52,156	+6%	+634,079	+82%	+9,187	**+1%**
	16	+3,261,957	+362%	-47,413	-5%	+1,887,958	+194%	+4,363	**+1%**
	19	+1,602,653	+154%	-62,820	-6%	+354,102	+29%	-4,402	**-1%**
	22	+2,690,624	+192%	-41,482	-3%	+1,283,918	+80%	+770	**0%**
	25	+1,891,115	+103%	-202,161	-11%	+1,925,530	+101%	-61,049	**-4%**
	30	+2,075,434	+84%	-415,567	-14%	+2,973,921	+115%	-76,920	**-3%**
	Best		0/10		0/10		0/10		10/10
	Max dev.	+5,218,581	1,768%	+839,366	303%	+1,019,063	351%	+41,078	5%

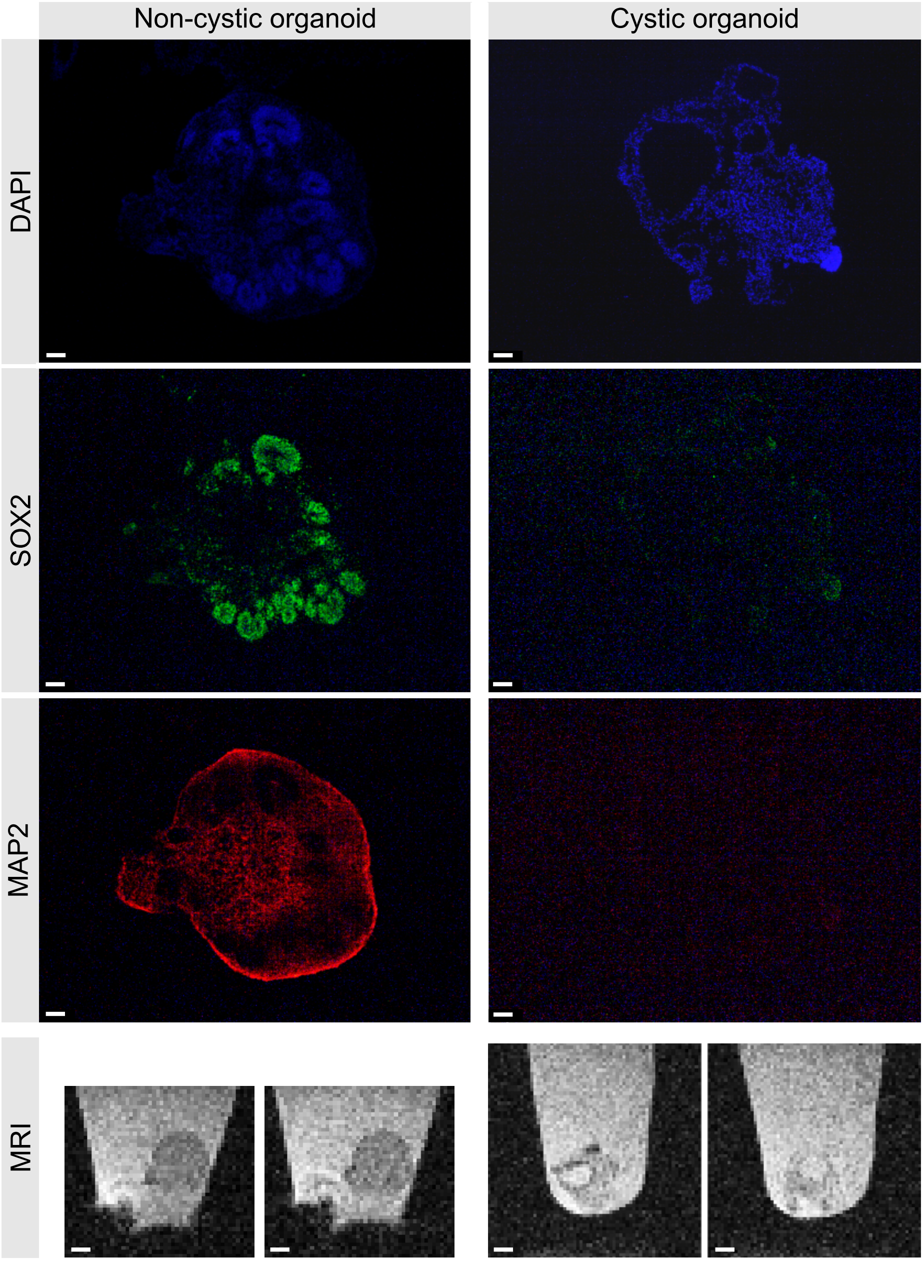


Figure A.6: Immunofluorescence stainings and MRI images of non-cystic and cystic organoid from Data$_B$. The non-cystic organoid grows more compact and shows ventricular-like structures containing SOX2+ progenitors and surrounding MAP2+ neurons while the cystic organoid only shows fluid-filled cavities (or "cysts") and no generation of neural progenitors and neurons. For better visibility, the brightness and contrast of the immunofluorescence images were enhanced. Scale bars: 50 μm (immunofluorescence images), 400 μm (MRI). Derivative of Fig. S4 by Deininger et al. [178] licensed under CC BY 4.0, with modified layout.

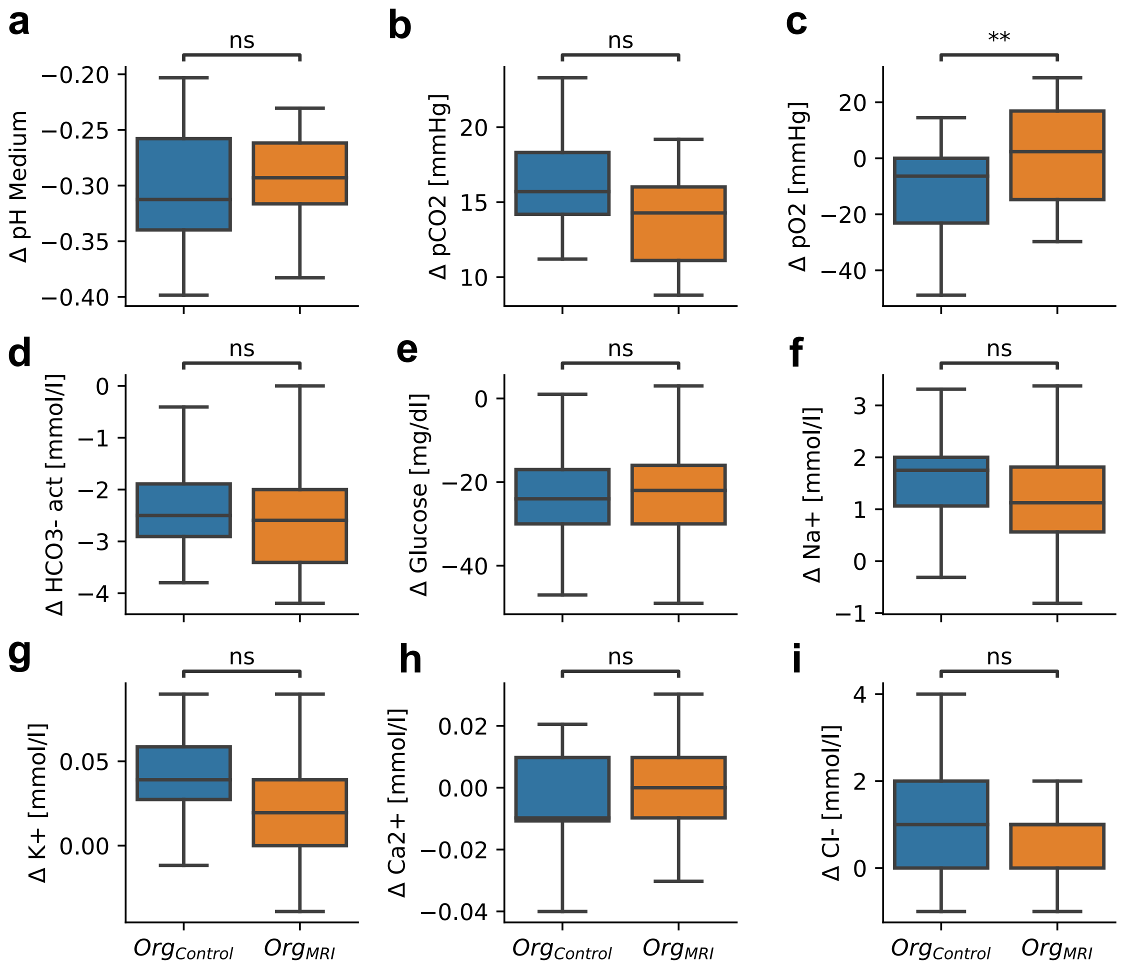


Figure A.7: Effect of MRI on organoids. Differences of all pre- and post-MRI measurements of medium from MRI organoids (Org_{MRI}), and control organoids ($Org_{control}$). Changes in pH of the medium (a), partial pressure of CO_2 (b), partial pressure of O_2 (c), bicarbonate (HCO_3^-) concentration (d), glucose levels (e), sodium (Na^+) concentration (f), potassium (K^+) concentration (g), calcium (Ca^{2+}) concentration (h), and chloride (Cl^-) concentration (i). Control organoids did not undergo MRI, but the same transportation to the MRI scanner. $^{**}p<0.01$, ns: not significant, t-test with Bonferroni correction for multiple hypothesis testing.

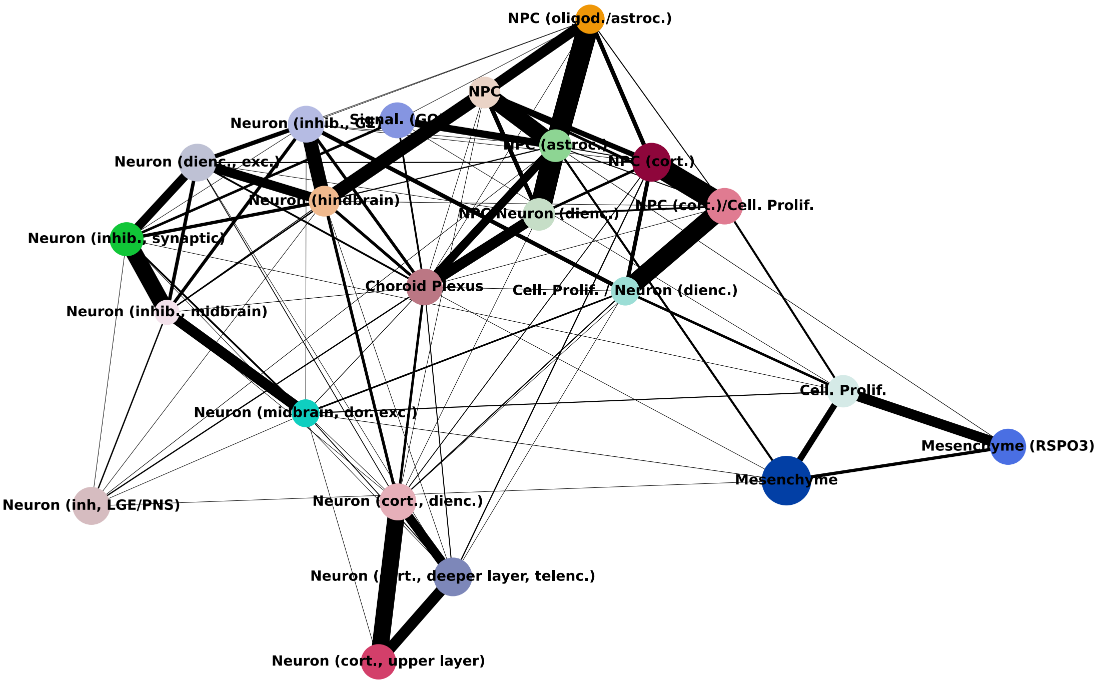


Figure A.8: PAGA graph illustrating a potential differentiation trajectory from proliferating cells through NPCs to multiple neuronal subtypes.

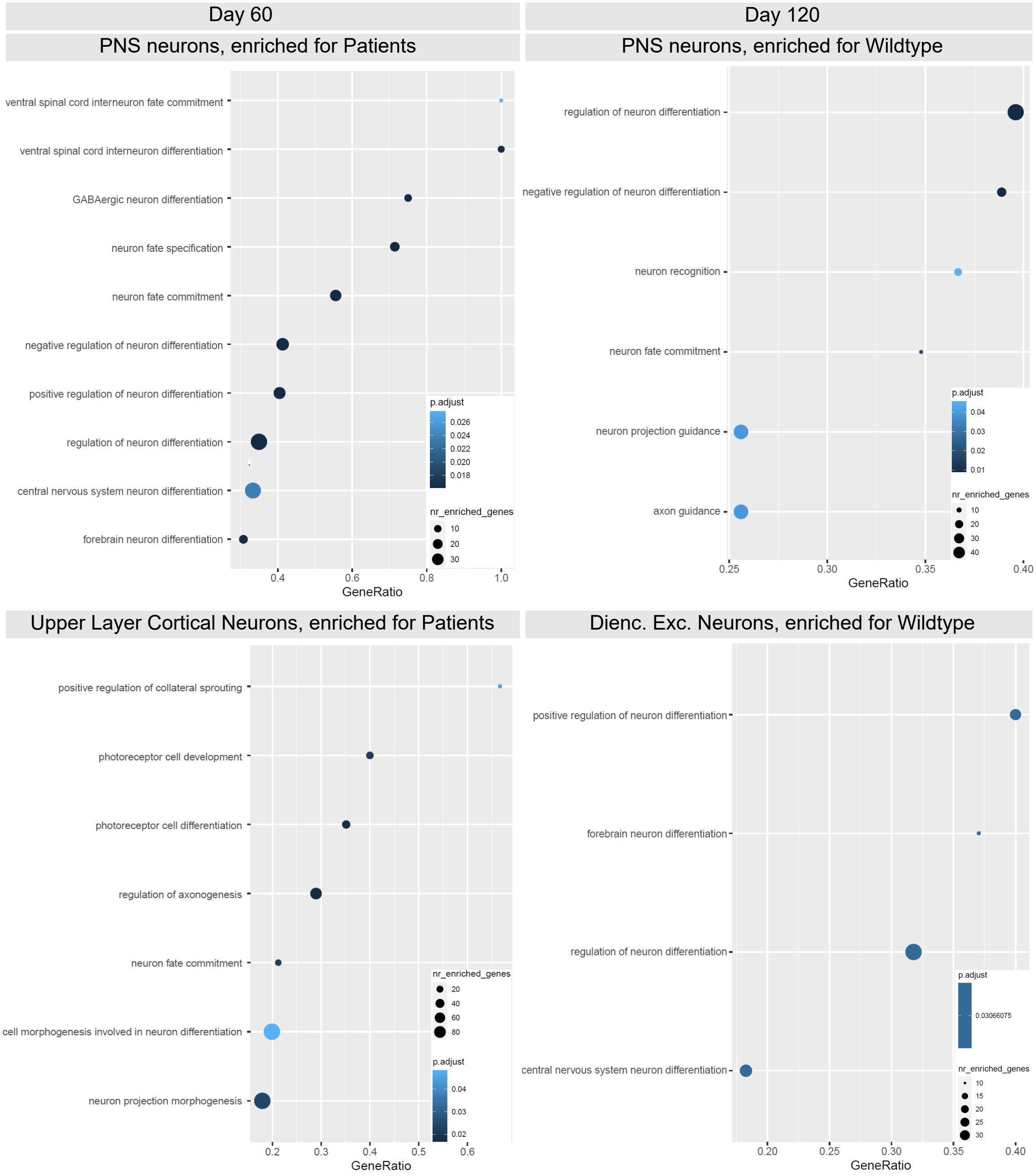


Figure A.9: Gene set enrichment analysis suggests earlier neuron differentiation in SSADH-Patients: additional examples. No neuron differentiation-related gene sets were enriched for Wildtype organoids on day 60, and no gene sets were enriched for Patient organoids on day 120.

A.2. Analysis Pipelines for ETiX-Embryos

Table A.3: Performance of different strategies for transferring synchronized time points from advanced-stage to early-stage ETiX-embryo classification. ResNet is the baseline model trained and evaluated on unsynchronized single time points. StembryoNet$_A$ applied a time shift of the synchronized time points. StembryoNet$_B$ and StembryoNet$_C$ combined shifting with shrinkage, assuming greater desynchronization at later stages and lower desynchronization at earlier stages. A linear transformation was applied to shift and shrink the original time range 65 h – 90 h to a target range (e.g. 35 h – 50 h), preserving the relative temporal spacing between time points. All models were trained using five times repeated 5-fold cross validation. Accuracy (ACC) and F1-score (F1) are reported as mean ± SD.

Classification	**Model**	**Time**	**ACC**	**F1**
Advanced-stage	StembryoNet	65h - 90h	0.88 ± 0.02	0.77 ± 0.03
Early-stage	ResNet	50h	0.74 ± 0.04	0.58 ± 0.04
Early-stage	StembryoNet$_A$	25h - 50h	0.76 ± 0.04	0.57 ± 0.05
Early-stage	StembryoNet$_B$	35h - 50h	0.77 ± 0.04	0.59 ± 0.05
Early-stage	StembryoNet$_C$	45h - 50h	0.78 ± 0.03	0.58 ± 0.06

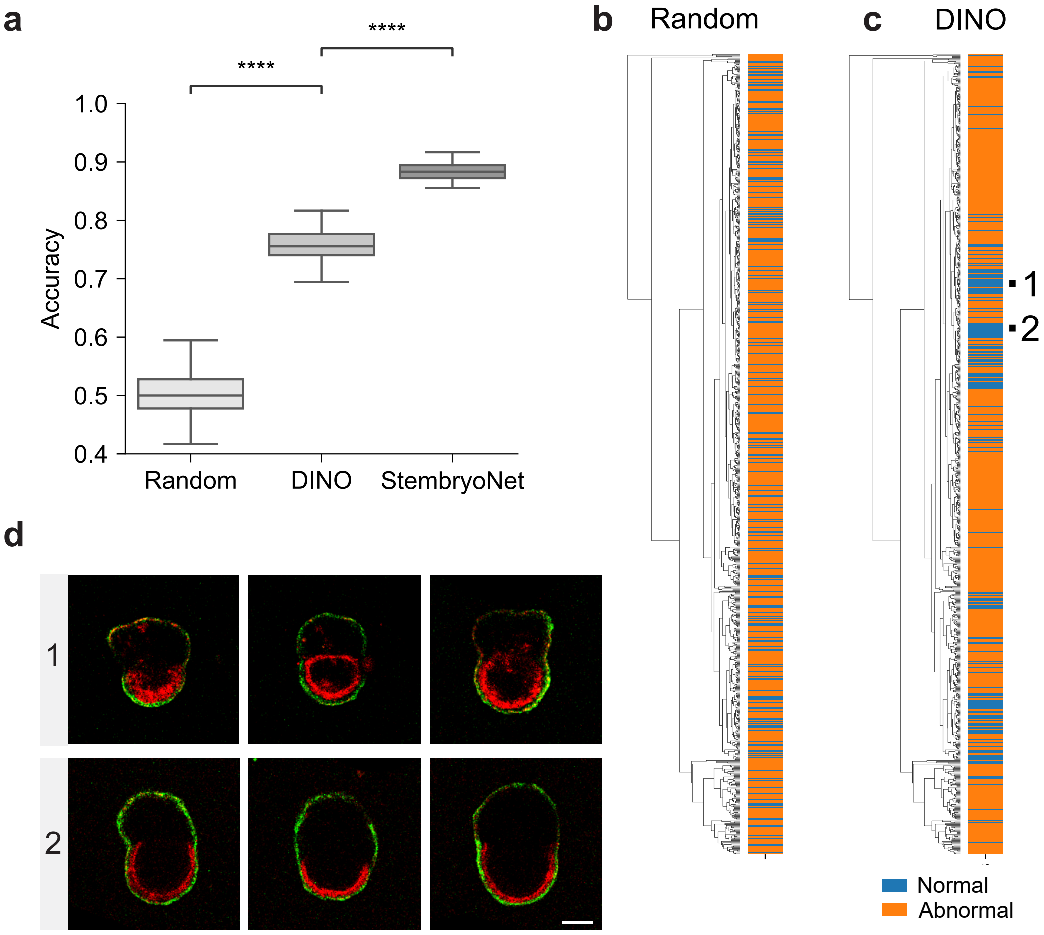


Figure A.10: Performance of self-supervised learning model and comparison with StembryoNet. (a) Performance comparison between an XGBoost model trained on DINO embeddings of ETiX-embryos and the StembryoNet model. The DINO model was trained on fluorescence in-focus embryo images captured between 65 and 90 hours. (b-c) Hierarchical clustering of DINO ETiX-embryo embeddings at the synchronized time point, as annotated by the expert embryologist. Colors represent the ground truth ETiX-embryo labels. In (b), class labels were randomly shuffled to demonstrate DINO's clustering performance compared to random clustering. (d) Selected samples of two sets of clustered normal embryos from (c). Scale bar: 100 μm. ****p<0.0001, two-sided Welch's t-test. Adapted from Fig. S4 in Caldarelli, Deininger et al. [40], licensed under CC BY 4.0, with modified layout and colors.

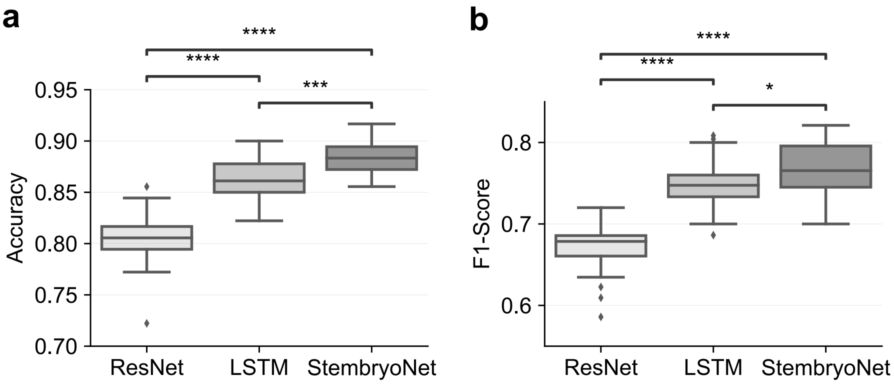


Figure A.11: StembryoNet outperforms LSTM-based approach. For LSTM-based ETiX-embryo classification, the trained StembryoNet backbone generates raw predictions on the training, validation, and test sets during the final 25 hours of ETiX-embryo development. These time-series predictions serve as input for the LSTM, which—like StembryoNet —is evaluated using five times repeated 5-fold cross-validation with identical data splits. The LSTM model architecture consisted of an input dimension of 2, a hidden dimension of 100, and an output dimension of 2. Training was performed with a batch size of 64 using the Adam optimizer with a learning rate of 0.001. A Cross-Entropy Loss weighted by inverse class frequencies was utilized to address class imbalance. The learning rate was scheduled using StepLR with a step size of 10 and a decay factor of 0.5. Training was conducted for 50 epochs, and the best-performing model was selected based on validation F1-score. *P<0.05, ***P<0.001, ****p<0.0001, two-sided Welch's t-test.

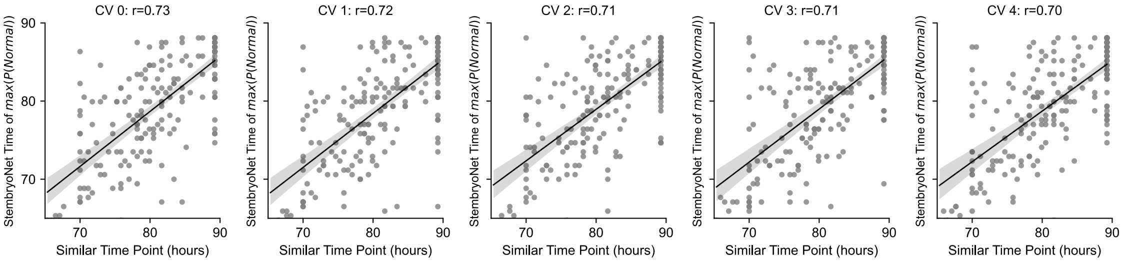


Figure A.12: StembryoNet's predicted time point of maximum probability for normal ETiX-embryo development correlates with the embryologist's annotated synchronized time point. Each subplot is based on StembryoNet test predictions across one cross validation split. Synchronization of abnormal ETiX-embryos was not feasible; their time points were randomly assigned, so only normal ETiX-embryos are shown. *r*, Pearson correlation coefficient. Error bar: CI, 95%.

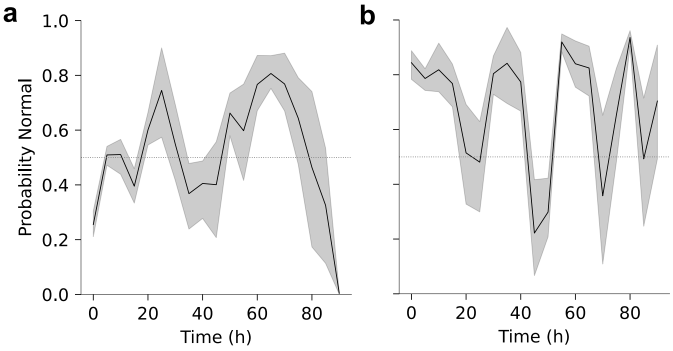


Figure A.13: ETiX-embryo prediction trajectories in the *Other* class. (a, b) ResNet predictions of the probability of being classified as normal for selected ETiX-embryos plotted over the entire observation period. For this application, ResNet$_{\text{Fluor}}$ was trained on unsynchronized data on distinct time points, from 0 to 90 hours in 5-hour intervals (Figure 3.5a,b), using five repeated 5-fold cross-validation runs (5×5-fold CV). (a) Embryo ds2_109 (abnormal) is assigned to the *Other* class due to contradicting slope (0.0017) and values at 0 hours (0.25) and 90 hours (0.04). The positive slope would indicate a transition from abnormal to normal, while the values at 0 and 90 hours indicate continuously abnormal class. (b) Embryo ds2_263 (normal) is also placed in the *Other* class due to contradicting slope (−0.0014) and values at 0 hours (0.86) and 90 hours (0.82). The negative slope would indicate a transition from normal to abnormal, while the values at 0 and 90 hours indicate a continuously normal ETiX-embryo.

Abbreviations and Symbols

This chapter provides an overview of the abbreviations (Section A.3) and symbols (Section A.4) used throughout the thesis.

A.3. Abbreviations

Abbreviation	Description
2D	Two-dimensional
3D	Three-dimensional
Adam	An optimization algorithm used in deep learning
AdamW	Adam optimizer with weight decay
AI	Artificial Intelligence
ALDH5A1	Aldehyde Dehydrogenase 5 Family Member A1
B2A-2	A patient-derived clone with tubulinopathy
BCE	Binary Cross-Entropy
BFM	Brightfield Microscopy
B-cells	A type of immune cell (also called B lymphocytes)
CellPose	A deep learning model for cell segmentation
CNN	Convolutional Neural Network
Compactness	A metric used to assess the cysticity of brain organoids
CV	Cross-Validation
DAPI	Nuclei marker in fluorescence microscopy
DGEA	Differential Gene Expression Analysis
Dice score	A metric used to evaluate image segmentation performance
DINO	Self-supervised Deep Learning model
DTI	Diffusion Tensor Imaging
EFM	Epifluorescence Microscopy
Edge ratio	A metric for evaluating organoid extension beyond image borders
ESC	Embryonic Stem Cells, occurring in ETiX-embryos
ESC-iGata4	ESC cells with the iGata4 marker, occurring in ETiX-embryos
ETiX-embryo	Specific organism-level stem cell-derived model
F1	F1-score
FN	False negatives
FP	False positives
GAP	Global average pooling

Abbreviation	Description
GeneRatio	Proportion of differentially expressed genes that belong to a specific GO term
GO	Gene Ontology
Grad-CAM	Gradient-weighted Class Activation Mapping
GUI	Graphical User Interface
GT	Ground Truth
GSEA	Gene Set Enrichment Analysis
IFC	Imaging Flow Cytometry
LI	Live Imaging
LOOCV	Leave-One-Out Cross-Validation
LR	Logistic Regression
LSTM	Long Short-Term Memory
μm	Micrometer
mm	Millimeter
MiT-B0	A configuration of a SegFormer
MLP	Multilayer Perceptron
MOrgAna	A deep learning-based organoid analysis tool
$\text{MOrgAna}_{\text{LR,C}}$	MOrgAna using logistic regression with classification masks
$\text{MOrgAna}_{\text{LR,W}}$	MOrgAna using logistic regression with watershed masks
$\text{MOrgAna}_{\text{MLP,C}}$	MOrgAna using MLP with classification masks
$\text{MOrgAna}_{\text{MLP,W}}$	MOrgAna using MLP with watershed masks
MRI	Magnetic Resonance Imaging
MViT	Multiscale Vision Transformer
NA	Not Applicable
NPCs	Neural Progenitor Cells
OrganoSeg	Organoid analysis tool
Otsu	Otsu's thresholding method
PAGA	Partition-based Graph Abstraction
PC1, PC2	First and second principal components of PCA
PCA	Principal Component Analysis
Pearson	Pearson correlation coefficient
PyRadiomics	A tool for extracting morphological features from images
px	Pixel
ResNet	Residual Network
RNN	Recurrent Neural Network
ROC AUC	Receiver Operating Characteristic Area Under the Curve
SAM	Segment Anything Model
scRNA-Seq	Single-cell RNA Sequencing
SD	Standard Deviation
SegFormer	A deep learning-based model for organoid segmentation and analysis
SN_{A1}, SN_{A2}, SN_{A3}	StembryoNet models trained on annotations A1-A3

Abbreviation	Description
SOX2	Marker for neural stem cells
SSADH	Succinic Semialdehyde Dehydrogenase
SVM	Support Vector Machine
T2*-w	T2*-weighted, MRI sequence
T-cells	A type of immune cell
TH2-7	A patient-derived clone with neurotransmitter disorder
TN	True negatives
TP	True positives
TSC	Trophoblast Stem Cells
UMAP	Uniform Manifold Approximation and Projection
U-Net	CNN for image segmentation
ViT	Vision Transformer
vx	Voxel
wt2D	A healthy control clone
XGBoost	Extreme Gradient Boosting
Z-score	Normalization method

A.4. Symbols

The symbols follow general rules outlined at the beginning of the section. The subsequent table lists all symbols used in the thesis, along with their descriptions, and is partly based on metrics introduced in [108].

- Lowercase letters (e.g., x, y, z, t, i, j) denote scalar variables, indices, and coordinates.
- Uppercase letters (e.g., M, A, B, C) denote images, masks, or sets.
- $|\cdot|$ denotes absolute value or set cardinality, depending on context.
- $\setminus$ indicates set difference.
- $\binom{a}{2}$ denotes the binomial coefficient, i.e., the number of unique unordered pairs among a elements.
- $\mu(X), \sigma(X)$ denote mean and standard deviation in the image, mask, or set X.
- $\min(X), \max(X)$ denote the minimum and maximum values in the image, mask, or set X.
- $\log(\cdot)$ denotes the natural logarithm.
- $\arg\max$ returns the argument that maximizes the given expression.

Symbol	Description
α, β	Weighting factors for $\mathrm{L}(y, \hat{y})$
$A_{\mathrm{org}}^{\mu\mathrm{m}^2}, A_{\mathrm{org}}^{\mathrm{px}^2}$	Organoid area, measured in $\mu\mathrm{m}^2$ or px^2
β_1, β_2	AdamW optimizer parameters
$\boldsymbol{C}$	Certainty vector containing C_i values for all images in the dataset
C_i	Certainty value for image i
fp_i	False positives: split or added nuclei
fn_i	False negatives: missed nuclei
K	Set of class labels for semantic segmentation or image classification
$\mathrm{L}(y, \hat{y})$	Combined loss: $\alpha \cdot \mathrm{L}_{\mathrm{BCE}} + \beta \cdot \mathrm{L}_{\mathrm{Dice}}$
$\mathrm{L}_{\mathrm{BCE}}$	Binary cross-entropy loss
$\mathrm{L}_{\mathrm{Dice}}$	Dice loss
$\mathcal{M}$	Set of mosaic tiles in live-imaging microscopy
$M_{\mathrm{A1}}, M_{\mathrm{A2}}, M_{\mathrm{A3}}$	Models trained on annotations from annotators A1–A3, respectively
M_{AC}	Model trained on the consensus annotation derived from A1–A3
$M_{i,j}$	Segmentation mask for image i by annotator/model j
$M(p)$	Segmentation mask value at spatial location p
$m(x, y, z, t)$	Intensity of mosaic tile m at position (x, y, z) and time point t
N	Total number of samples in dataset
$N_{\mathrm{split}}, N_{\mathrm{miss}}, N_{\mathrm{add}}$	Number of split, missed, and added nuclei errors
P	Set of all pairwise annotation or prediction combinations
p	Spatial location in a 2D image ($p = (x, y)$) or 3D image ($p = (x, y, z)$)
p	p-value from statistical hypothesis testing
$P(k \mid p)$	Predicted probability of class k at location p
Q_{CC}	Sample quality based on cell count correlation across replicates
Q_F	Quality metric for nuclei segmentation
Q_{MCR}	Sample quality based on mesenchymal cell ratio across replicates
$Q_{P,i}, Q_{R,i}$	Precision-like and recall-like values for image i
r	Pearson correlation coefficient (Pearson's r)
$s_{\mu\mathrm{m}/\mathrm{px}}$	Microscopy scaling factor for unit conversion
θ	Threshold for StembryoNet-based ETiX-embryo classification
tp_i	Adjusted true positives for image i
$\mathbf{v}_{\mathrm{medium}}$	Intensities of medium voxels
$\mathbf{v}_{\mathrm{org}}$	Intensities of organoid voxels
$X_{\mathrm{tile}}, Y_{\mathrm{tile}}$	Coordinate sets for a single tile (768 × 768)
$X_{\mathrm{mos}}, Y_{\mathrm{mos}}$	Coordinate sets for full mosaic image (3840 × 3840)
$X_{\mathrm{emb}}, Y_{\mathrm{emb}}$	Coordinate sets of extracted ETiX-embryo image (288 × 288)
$y, \hat{y}$	Ground truth and predicted labels or masks

List of Own Publications

[1] **L. Deininger***, P. Caldarelli*, M. Zernicka-Goetz, et al. *From pixels to patterns: The AI revolution in stem cell-derived models.* Submitted to Nature Methods. 2025.

[178] **L. Deininger***, S. Jung-Klawitter*, R. Mikut, et al. *An AI-based segmentation and analysis pipeline for high-field MR monitoring of cerebral organoids.* In: *Scientific Reports*, Vol. 13, No. 1 (Dec. 2023). ISSN: 2045-2322. DOI: 10.1038/s41598-023-48343-7.

[198] J. Schröter*, **L. Deininger***, B. Lupse, et al. *A large and diverse brain organoid dataset of 1,400 cross-laboratory images of 64 trackable brain organoids.* In: *Scientific Data*, Vol. 11, No. 1 (May 2024). ISSN: 2052-4463. DOI: 10.1038/s41597-024-03330-z.

[215] D. Vonficht*, L. Jopp-Saile*, S. Yousefian*, . . . , **L. Deininger**, et al. *Ultra-high scale cytometry-based cellular interaction mapping.* Accepted at Nature Methods. 2025.

[40] P. Caldarelli*, **L. Deininger***, S. Zhao, et al. *AI-based approach to dissect the variability of mouse stem cell-derived embryo models.* In: *Nature Communications*, Vol. 16, No. 1 (Feb. 2025). ISSN: 2041-1723. DOI: 10.1038/s41467-025-56908-5.

* Shared first authorship.

Bibliography

[1] L. Deininger, P. Caldarelli, M. Zernicka-Goetz, and R. Mikut. *From pixels to patterns: The AI revolution in stem cell-derived models.* Submitted to Nature Methods. 2025.

[2] M. A. Lancaster, M. Renner, C.-A. Martin, et al. *Cerebral organoids model human brain development and microcephaly.* In: *Nature*, Vol. 501, No. 7467 (2013), 10.1038/nature12517. ISSN: 0028-0836. DOI: `10.1038/nature12517`.

[3] H. Hu, H. Gehart, B. Artegiani, et al. *Long-Term Expansion of Functional Mouse and Human Hepatocytes as 3D Organoids.* In: *Cell*, Vol. 175, No. 6 (2018), 1591–1606.e19. ISSN: 0092-8674. DOI: `10.1016/j.cell.2018.11.013`.

[4] T. Sato, R. G. Vries, H. J. Snippert, et al. *Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.* In: *Nature*, Vol. 459, No. 7244 (2009), pp. 262–265. ISSN: 0028-0836. DOI: `10.1038/nature07935`.

[5] W. M. S. Russell and R. L. Burch. *The principles of humane experimental technique.* London: Methuen and Co. Ltd., 1959, xiv + 238 pp.

[6] J. G. Camp, F. Badsha, M. Florio, et al. *Human cerebral organoids recapitulate gene expression programs of fetal neocortex development.* In: *Proceedings of the National Academy of Sciences*, Vol. 112, No. 51 (Dec. 2015), pp. 15672–15677. ISSN: 1091-6490. DOI: `10.1073/pnas.1520760112`.

[7] M. Takasato, P. X. Er, H. S. Chiu, et al. *Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis.* In: *Nature*, Vol. 526, No. 7574 (Oct. 2015), pp. 564–568. ISSN: 1476-4687. DOI: `10.1038/nature15695`.

[8] T. Takebe, K. Sekine, M. Enomura, et al. *Vascularized and functional human liver from an iPSC-derived organ bud transplant.* In: *Nature*, Vol. 499, No. 7459 (July 2013), pp. 481–484. ISSN: 1476-4687. DOI: `10.1038/nature12271`.

[9] J. R. Spence, C. N. Mayhew, S. A. Rankin, et al. *Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro.* In: *Nature*, Vol. 470, No. 7332 (Dec. 2010), pp. 105–109. ISSN: 1476-4687. DOI: `10.1038/nature09691`.

[10] P. P. Garcez, E. C. Loiola, R. Madeiro da Costa, et al. *Zika virus impairs growth in human neurospheres and brain organoids.* In: *Science*, Vol. 352, No. 6287 (May 2016), pp. 816–818. ISSN: 1095-9203. DOI: `10.1126/science.aaf6116`.

[11] H. Kim, H. J. Park, H. Choi, et al. *Modeling G2019S-LRRK2 Sporadic Parkinson's Disease in 3D Midbrain Organoids.* In: *Stem Cell Reports*, Vol. 12, No. 3 (Mar. 2019), pp. 518–531. ISSN: 2213-6711. DOI: `10.1016/j.stemcr.2019.01.020`.

[12] C. Chen, P. Jiang, H. Xue, et al. *Role of astroglia in Down's syndrome revealed by patient-derived human-induced pluripotent stem cells.* In: *Nature Communications*, Vol. 5, No. 1 (July 2014). ISSN: 2041-1723. DOI: `10.1038/ncomms5430`.

[13] M. A. Israel, S. H. Yuan, C. Bardy, et al. *Probing sporadic and familial Alzheimer's disease using induced pluripotent stem cells.* In: *Nature*, Vol. 482, No. 7384 (Jan. 2012), pp. 216–220. ISSN: 1476-4687. DOI: `10.1038/nature10821`.

[14] M. F. Burkhardt, F. J. Martinez, S. Wright, et al. *A cellular model for sporadic ALS using patient-derived induced pluripotent stem cells.* In: *Molecular and Cellular Neuroscience*, Vol. 56 (Sept. 2013), pp. 355–364. ISSN: 1044-7431. DOI: `10.1016/j.mcn.2013.07.007`.

[15] K. J. Brennand, A. Simone, J. Jou, et al. *Modelling schizophrenia using human induced pluripotent stem cells.* In: *Nature*, Vol. 473, No. 7346 (Apr. 2011), pp. 221–225. ISSN: 1476-4687. DOI: `10.1038/nature09915`.

[16] G. Vlachogiannis, S. Hedayat, A. Vatsiou, et al. *Patient-derived organoids model treatment response of metastatic gastrointestinal cancers.* In: *Science*, Vol. 359, No. 6378 (Feb. 2018), pp. 920–926. ISSN: 1095-9203. DOI: `10.1126/science.aao2774`.

[17] N. D. Germain, P.-F. Chen, A. M. Plocik, et al. *Gene expression analysis of human induced pluripotent stem cell-derived neurons carrying copy number variants of chromosome 15q11-q13.1.* In: *Molecular Autism*, Vol. 5, No. 1 (Aug. 2014). ISSN: 2040-2392. DOI: `10.1186/2040-2392-5-44`.

[18] J. M. Madison, F. Zhou, A. Nigam, et al. *Characterization of bipolar disorder patient-specific induced pluripotent stem cells from a family reveals neurodevelopmental and mRNA expression abnormalities.* In: *Molecular Psychiatry*, Vol. 20, No. 6 (Mar. 2015), pp. 703–717. ISSN: 1476-5578. DOI: `10.1038/mp.2015.7`.

[19] M. Igoillo-Esteve, E. Gurgul-Convey, A. Hu, et al. *Unveiling a common mechanism of apoptosis in beta-cells and neurons in Friedreich's ataxia.* In: *Human Molecular Genetics*, Vol. 24, No. 8 (Dec. 2014), pp. 2274–2286. ISSN: 0964-6906. DOI: `10.1093/hmg/ddu745`.

[20] A. Leung, S. K. Nah, W. Reid, et al. *Induced Pluripotent Stem Cell Modeling of Multisystemic, Hereditary Transthyretin Amyloidosis.* In: *Stem Cell Reports*, Vol. 1, No. 5 (Nov. 2013), pp. 451–463. ISSN: 2213-6711. DOI: `10.1016/j.stemcr.2013.10.003`.

[21] N. Sachs, J. DE Ligt, O. Kopper, et al. *A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity.* In: *Cell*, Vol. 172, No. 1–2 (Jan. 2018), 373–386.e10. ISSN: 0092-8674. DOI: `10.1016/j.cell.2017.11.010`.

[22] S. N. Ooft, F. Weeber, K. K. Dijkstra, et al. *Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients.* In: *Science Translational Medicine*, Vol. 11, No. 513 (Oct. 2019). ISSN: 1946-6242. DOI: `10.1126/scitranslmed.aay2574`.

[23] X. Qian, H. N. Nguyen, M. M. Song, et al. *Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure.* In: *Cell*, Vol. 165, No. 5 (May 2016), pp. 1238–1254. ISSN: 0092-8674. DOI: `10.1016/j.cell.2016.04.032`.

[24] G. Quadrato, T. Nguyen, E. Z. Macosko, et al. *Cell diversity and network dynamics in photosensitive human brain organoids.* In: *Nature*, Vol. 545, No. 7652 (Apr. 2017), pp. 48–53. ISSN: 1476-4687. DOI: `10.1038/nature22047`.

[25] M. Hofer and M. P. Lutolf. *Engineering organoids.* In: *Nature Reviews Materials*, Vol. 6, No. 5 (Feb. 2021), pp. 402–420. ISSN: 2058-8437. DOI: `10.1038/s41578-021-00279-y`.

[26] K. Y. C. Lau, G. Amadei, and M. Zernicka-Goetz. *Assembly of complete mouse embryo models from embryonic and induced stem cell types in vitro.* In: *Nature Protocols*, Vol. 18, No. 12 (Oct. 2023), pp. 3662–3689. ISSN: 1750-2799. DOI: `10.1038/s41596-023-00891-y`.

[27] A. Geidies, M. L. Medar, and H. M. Beyer. *Engineering organoids as cerebral disease models.* In: *Current Opinion in Biotechnology*, Vol. 92 (Apr. 2025), p. 103253. ISSN: 0958-1669. DOI: `10.1016/j.copbio.2024.103253`.

[28] C. Jensen and Y. Teng. *Is It Time to Start Transitioning From 2D to 3D Cell Culture?* In: *Frontiers in Molecular Biosciences*, Vol. 7 (Mar. 2020). ISSN: 2296-889X. DOI: `10.3389/fmolb.2020.00033`.

[29] R. Edmondson, J. J. Broglie, A. F. Adcock, and L. Yang. *Three-Dimensional Cell Culture Systems and Their Applications in Drug Discovery and Cell-Based Biosensors.* In: *ASSAY and Drug Development Technologies*, Vol. 12, No. 4 (May 2014), pp. 207–218. ISSN: 1557-8127. DOI: `10.1089/adt.2014.573`.

[30] K. Duval, H. Grover, L.-H. Han, et al. *Modeling Physiological Events in 2D vs. 3D Cell Culture.* In: *Physiology*, Vol. 32, No. 4 (July 2017), pp. 266–277. ISSN: 1548-9221. DOI: `10.1152/physiol.00036.2016`.

[31] M. A. Lancaster and J. A. Knoblich. *Generation of cerebral organoids from human pluripotent stem cells.* In: *Nature Protocols*, Vol. 9, No. 10 (Sept. 2014), pp. 2329–2340. ISSN: 1750-2799. DOI: `10.1038/nprot.2014.158`.

[32] J. Kim, B.-K. Koo, and J. A. Knoblich. *Human organoids: model systems for human biology and medicine.* In: *Nature Reviews Molecular Cell Biology*, Vol. 21, No. 10 (2020), pp. 571–584. ISSN: 1471-0072. DOI: `10.1038/s41580-020-0259-3`.

[33] J. Badai, Q. Bu, and L. Zhang. *Review of Artificial Intelligence Applications and Algorithms for Brain Organoid Research.* In: *Interdisciplinary Sciences: Computational Life Sciences*, Vol. 12, No. 4 (2020), pp. 383–394. ISSN: 1913-2751. DOI: `10.1007/s12539-020-00386-4`.

[34] W. K. Raja, A. E. Mungenast, Y.-T. Lin, et al. *Self-Organizing 3D Human Neural Tissue Derived from Induced Pluripotent Stem Cells Recapitulate Alzheimer's Disease Phenotypes.* In: *PLoS ONE*, Vol. 11, No. 9 (2016), e0161969. DOI: `10.1371/journal.pone.0161969`.

[35] H. Kim, H. J. Park, H. Choi, et al. *Modeling G2019S-LRRK2 Sporadic Parkinson's Disease in 3D Midbrain Organoids.* In: *Stem Cell Reports*, Vol. 12, No. 3 (2019), pp. 518–531. ISSN: 2213-6711. DOI: `10.1016/j.stemcr.2019.01.020`.

[36] S. T. Foliaki, B. Schwarz, B. R. Groveman, et al. *Neuronal excitatory-to-inhibitory balance is altered in cerebral organoid models of genetic neurological diseases.* In: *Molecular Brain*, Vol. 14, No. 1 (2021), p. 156. DOI: `10.1186/s13041-021-00864-w`.

[37] M. Bershteyn, T. J. Nowakowski, A. A. Pollen, et al. *Human iPSC-Derived Cerebral Organoids Model Cellular Features of Lissencephaly and Reveal Prolonged Mitosis of Outer Radial Glia.* In: *Cell Stem Cell*, Vol. 20, No. 4 (Apr. 2017), 435–449.e4. ISSN: 1934-5909. DOI: `10.1016/j.stem.2016.12.007`.

[38] H. Wang. *Modeling Neurological Diseases With Human Brain Organoids.* In: *Frontiers in Synaptic Neuroscience*, Vol. 10 (June 2018). ISSN: 1663-3563. DOI: `10.3389/fnsyn.2018.00015`.

[39] K. Sabitha, A. K. Shetty, and D. Upadhya. *Patient-derived iPSC modeling of rare neurodevelopmental disorders: Molecular pathophysiology and prospective therapies.* In: *Neuroscience; Biobehavioral Reviews*, Vol. 121 (Feb. 2021), pp. 201–219. ISSN: 0149-7634. DOI: `10.1016/j.neubiorev.2020.12.025`.

[40] P. Caldarelli, L. Deininger, S. Zhao, et al. *AI-based approach to dissect the variability of mouse stem cell-derived embryo models.* In: *Nature Communications*, Vol. 16, No. 1 (Feb. 2025). ISSN: 2041-1723. DOI: `10.1038/s41467-025-56908-5`.

[41] X. Qian, H. Song, and G.-l. Ming. *Brain organoids: advances, applications and challenges.* In: *Development*, Vol. 146, No. 8 (Apr. 2019). ISSN: 0950-1991. DOI: `10.1242/dev.166074`.

[42] C. N. Mayhew and R. Singhania. *A review of protocols for brain organoids and applications for disease modeling.* In: *STAR Protocols*, Vol. 4, No. 1 (Mar. 2023), p. 101860. ISSN: 2666-1667. DOI: `10.1016/j.xpro.2022.101860`.

[43] F. Kaluthantrige Don and N. Kalebic. *Forebrain Organoids to Model the Cell Biology of Basal Radial Glia in Neurodevelopmental Disorders and Brain Evolution.* In: *Frontiers in Cell and Developmental Biology*, Vol. 10 (June 2022). ISSN: 2296-634X. DOI: `10.3389/fcell.2022.917166`.

[44] D. Lindenhofer, S. Haendeler, C. Esk, et al. *Cerebral organoids display dynamic clonal growth and tunable tissue replenishment.* In: *Nature Cell Biology*, Vol. 26, No. 5 (May 2024), pp. 710–718. ISSN: 1476-4679. DOI: `10.1038/s41556-024-01412-z`.

[45] M. A. Lancaster, M. Renner, C.-A. Martin, et al. *Cerebral organoids model human brain development and microcephaly.* In: *Nature*, Vol. 501, No. 7467 (Aug. 2013), pp. 373–379. ISSN: 1476-4687. DOI: `10.1038/nature12517`.

[46] R. Guerrini and W. B. Dobyns. *Malformations of cortical development: clinical features and genetic causes.* In: *The Lancet Neurology*, Vol. 13, No. 7 (July 2014), pp. 710–726. ISSN: 1474-4422. DOI: `10.1016/s1474-4422(14)70040-7`.

[47] V. Iefremova, G. Manikakis, O. Krefft, et al. *An Organoid-Based Model of Cortical Development Identifies Non-Cell-Autonomous Defects in Wnt Signaling Contributing to Miller-Dieker Syndrome.* In: *Cell Reports*, Vol. 19, No. 1 (Apr. 2017), pp. 50–59. ISSN: 2211-1247. DOI: `10.1016/j.celrep.2017.03.047`.

[48] M. Didiasova, A. Banning, H. Brennenstuhl, et al. *Succinic Semialdehyde Dehydrogenase Deficiency: An Update.* In: *Cells*, Vol. 9, No. 2 (Feb. 2020), p. 477. ISSN: 2073-4409. DOI: `10.3390/cells9020477`.

[49] W. K. Raja, A. E. Mungenast, Y.-T. Lin, et al. *Self-Organizing 3D Human Neural Tissue Derived from Induced Pluripotent Stem Cells Recapitulate Alzheimer's Disease Phenotypes.* In: *PLOS ONE*, Vol. 11, No. 9 (Sept. 2016), e0161969. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0161969`.

[50] S. T. Foliaki, B. Schwarz, B. R. Groveman, et al. *Neuronal excitatory-to-inhibitory balance is altered in cerebral organoid models of genetic neurological diseases.* In: *Molecular Brain*, Vol. 14, No. 1 (Oct. 2021). ISSN: 1756-6606. DOI: `10.1186/s13041-021-00864-w`.

[51] E. Di Lullo and A. R. Kriegstein. *The use of brain organoids to investigate neural development and disease.* In: *Nature Reviews Neuroscience*, Vol. 18, No. 10 (Sept. 2017), pp. 573–584. ISSN: 1471-0048. DOI: `10.1038/nrn.2017.107`.

[52] H. Brennenstuhl, T. Opladen, D. Ebrahimi-Fakhari, et al. *The use of patient-derived induced pluripotent stem cells (IPSCS) and IPSC-derived cerebral organoids to explore pathomechanisms of succinic semialdehyde dehydrogenase deficiency.* In: *Abstracts of the 33rd Annual Conference of Association for pediatric metabolic disorders* (2019), p. 365. DOI: `10.1007/s00112-019-0665-9`.

[53] G. Quadrato, J. Brown, and P. Arlotta. *The promises and challenges of human brain organoids as models of neuropsychiatric disease.* In: *Nature Medicine*, Vol. 22, No. 11 (2016), pp. 1220–1228. ISSN: 1078-8956. DOI: `10.1038/nm.4214`.

[54] G. Quadrato, T. Nguyen, E. Z. Macosko, et al. *Cell diversity and network dynamics in photosensitive human brain organoids.* In: *Nature*, Vol. 545, No. 7652 (2017), pp. 48–53. ISSN: 0028-0836. DOI: `10.1038/nature22047`.

[55] X. Qian, H. Song, and G.-l. Ming. *Brain organoids: advances, applications and challenges.* In: *Development*, Vol. 146, No. 8 (2019), dev166074. ISSN: 0950-1991. DOI: `10.1242/dev.166074`.

[56] M. A. Lancaster and J. A. Knoblich. *Generation of cerebral organoids from human pluripotent stem cells.* In: *Nature Protocols*, Vol. 9, No. 10 (2014), pp. 2329–2340. ISSN: 1754-2189. DOI: `10.1038/nprot.2014.158`.

[57] M. Hofer and M. P. Lutolf. *Engineering organoids.* In: *Nature Reviews Materials*, Vol. 6, No. 5 (2021), pp. 402–420. ISSN: 2058-8437. DOI: `10.1038/s41578-021-00279-y`.

[58] G. Amadei, C. E. Handford, C. Qiu, et al. *Embryo model completes gastrulation to neurulation and organogenesis.* In: *Nature*, Vol. 610, No. 7930 (Aug. 2022), pp. 143–153. ISSN: 1476-4687. DOI: `10.1038/s41586-022-05246-3`.

[59] C. B. Martin, C. S. Chane, C. Clouchoux, and A. Histace. *Recent Trends and Perspectives in Cerebral Organoids Imaging and Analysis.* In: *Frontiers in Neuroscience*, Vol. 15 (2021), p. 629067. ISSN: 1662-4548. DOI: `10.3389/fnins.2021.629067`.

[60] S. M. Hickey, B. Ung, C. Bader, et al. *Fluorescence Microscopy—An Outline of Hardware, Biological Handling, and Fluorophore Considerations.* In: *Cells*, Vol. 11, No. 1 (Dec. 2021), p. 35. ISSN: 2073-4409. DOI: `10.3390/cells11010035`.

[61] C. Brémond Martin, C. Simon Chane, C. Clouchoux, and A. Histace. *Recent Trends and Perspectives in Cerebral Organoids Imaging and Analysis.* In: *Frontiers in Neuroscience*, Vol. 15 (July 2021). ISSN: 1662-453X. DOI: `10.3389/fnins.2021.629067`.

[62] Z. Ao, H. Cai, D. J. Havert, et al. *One-Stop Microfluidic Assembly of Human Brain Organoids To Model Prenatal Cannabis Exposure.* In: *Analytical Chemistry*, Vol. 92, No. 6 (Feb. 2020), pp. 4630–4638. ISSN: 1520-6882. DOI: `10.1021/acs.analchem.0c00205`.

[63] Y. Kang, Y. Zhou, Y. Li, et al. *A human forebrain organoid model of fragile X syndrome exhibits altered neurogenesis and highlights new treatment strategies.* In: *Nature Neuroscience*, Vol. 24, No. 10 (Aug. 2021), pp. 1377–1391. ISSN: 1546-1726. DOI: `10.1038/s41593-021-00913-6`.

[64] A. Albanese, J. M. Swaney, D. H. Yun, et al. *Multiscale 3D phenotyping of human cerebral organoids.* In: *Scientific Reports*, Vol. 10, No. 1 (2020), p. 21487. DOI: `10.1038/s41598-020-78130-7`.

[65] T. A. Shnaider and I. E. Pristyazhnyuk. *CLARITY and Light-Sheet microscopy sample preparation in application to human cerebral organoids.* In: *Vavilov Journal of Genetics and Breeding*, Vol. 25, No. 8 (Jan. 2022), pp. 889–895. ISSN: 2500-0462. DOI: `10.18699/vj21.103`.

[66] J. F. Dekkers, M. Alieva, L. M. Wellens, et al. *High-resolution 3D imaging of fixed and cleared organoids.* In: *Nature Protocols*, Vol. 14, No. 6 (May 2019), pp. 1756–1771. ISSN: 1750-2799. DOI: `10.1038/s41596-019-0160-8`.

[67] J. E. Rodriguez-Gatica, V. Iefremova, L. Sokhranyaeva, et al. *Imaging three-dimensional brain organoid architecture from meso- to nanoscale across development.* In: *Development*, Vol. 149, No. 20 (Aug. 2022). ISSN: 1477-9129. DOI: `10.1242/dev.200439`.

[68] K. Kartasalo, L. Latonen, J. Vihinen, et al. *Comparative analysis of tissue reconstruction algorithms for 3D histology.* In: *Bioinformatics*, Vol. 34, No. 17 (Apr. 2018), pp. 3013–3021. ISSN: 1367-4811. DOI: `10.1093/bioinformatics/bty210`.

[69] I. Wacker, W. Spomer, A. Hofmann, et al. *Hierarchical imaging: a new concept for targeted imaging of large volumes from cells to tissues.* In: *BMC Cell Biology*, Vol. 17, No. 1 (Dec. 2016). ISSN: 1471-2121. DOI: `10.1186/s12860-016-0122-8`.

[70] I. Wacker, W. Spomer, A. Hofmann, et al. *On the road to large volumes in LM and SEM: New tools for Array Tomography.* In: *Microscopy and Microanalysis*, Vol. 21, No. S3 (Aug. 2015), pp. 539–540. ISSN: 1435-8115. DOI: `10.1017/s1431927615003499`.

[71] K. Thorn. *A quick guide to light microscopy in cell biology.* In: *Molecular Biology of the Cell*, Vol. 27, No. 2 (Jan. 2016), pp. 219–222. ISSN: 1939-4586. DOI: `10.1091/mbc.e15-02-0088`.

[72] J. Wang, P. A. Gleeson, and L. Fourriere. *Long-term live cell imaging during differentiation of human iPSC-derived neurons.* In: *STAR Protocols*, Vol. 4, No. 4 (Dec. 2023), p. 102699. ISSN: 2666-1667. DOI: `10.1016/j.xpro.2023.102699`.

[73] M. Hesse, A. Raulf, G.-A. Pilz, et al. *Direct visualization of cell division using high-resolution imaging of M-phase of the cell cycle.* In: *Nature Communications*, Vol. 3, No. 1 (Sept. 2012). ISSN: 2041-1723. DOI: `10.1038/ncomms2089`.

[74] P. A. Santi. *Light Sheet Fluorescence Microscopy: A Review.* In: *Journal of Histochemistry and Cytochemistry*, Vol. 59, No. 2 (Feb. 2011), pp. 129–138. ISSN: 1551-5044. DOI: `10.1369/0022155410394857`.

[75] M. A. Sahraian and A. Eshaghi. *Role of MRI in diagnosis and treatment of multiple sclerosis.* In: *Clinical Neurology and Neurosurgery*, Vol. 112, No. 7 (Sept. 2010), pp. 609–615. ISSN: 0303-8467. DOI: `10.1016/j.clineuro.2010.03.022`.

[76] P. Mahlknecht, A. Hotter, A. Hussl, et al. *Significance of MRI in Diagnosis and Differential Diagnosis of Parkinson's Disease.* In: *Neurodegenerative Diseases*, Vol. 7, No. 5 (2010), pp. 300–318. ISSN: 1660-2862. DOI: `10.1159/000314495`.

[77] A. M. Al-Radaideh and E. M. Rababah. *The role of magnetic resonance imaging in the diagnosis of Parkinson's disease: a review.* In: *Clinical Imaging*, Vol. 40, No. 5 (Sept. 2016), pp. 987–996. ISSN: 0899-7071. DOI: `10.1016/j.clinimag.2016.05.006`.

[78] A. Beghin, G. Grenci, G. Sahni, et al. *Automated high-speed 3D imaging of organoid cultures with multi-scale phenotypic quantification.* In: *Nature Methods* (2022), pp. 1–12. ISSN: 1548-7091. DOI: `10.1038/s41592-022-01508-0`.

[79] L. Pellegrini, C. Bonfio, J. Chadwick, et al. *Human CNS barrier-forming organoids with cerebrospinal fluid production.* In: *Science*, Vol. 369, No. 6500 (2020). ISSN: 0036-8075. DOI: `10.1126/science.aaz5626`.

[80] J. Jo, Y. Xiao, A. X. Sun, et al. *Midbrain-like Organoids from Human Pluripotent Stem Cells Contain Functional Dopaminergic and Neuromelanin-Producing Neurons.* In: *Cell Stem Cell*, Vol. 19, No. 2 (Aug. 2016), pp. 248–257. ISSN: 1934-5909. DOI: `10.1016/j.stem.2016.07.005`.

[81] Y. Zhu, R. Huang, Z. Wu, et al. *Deep learning-based predictive identification of neural stem cell differentiation.* In: *Nature Communications*, Vol. 12, No. 1 (May 2021). ISSN: 2041-1723. DOI: `10.1038/s41467-021-22758-0`.

[82] R. Mukhopadhyay, P. Chandel, K. Prasad, and U. Chakraborty. *Machine learning aided single cell image analysis improves understanding of morphometric heterogeneity of human mesenchymal stem cells.* In: *Methods*, Vol. 225 (May 2024), pp. 62–73. ISSN: 1046-2023. DOI: `10.1016/j.ymeth.2024.03.005`.

[83] M. I. Love, W. Huber, and S. Anders. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* In: *Genome Biology*, Vol. 15, No. 12 (Dec. 2014). ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8.

[84] Y. Hao, T. Stuart, M. H. Kowalski, et al. *Dictionary learning for integrative, multimodal and scalable single-cell analysis.* In: *Nature Biotechnology* (2023). DOI: 10.1038/s41587-023-01767-y.

[85] M. Ashburner, C. A. Ball, J. A. Blake, et al. *Gene Ontology: tool for the unification of biology.* In: *Nature Genetics*, Vol. 25, No. 1 (May 2000), pp. 25–29. ISSN: 1546-1718. DOI: 10.1038/75556.

[86] A. Bazille, M. A. Guttman, E. R. McVeigh, and E. A. Zerhouni. *Impact of Semiautomated versus Manual Image Segmentation Errors on Myocardial Strain Calculation by Magnetic Resonance Tagging.* In: *Investigative Radiology*, Vol. 29, No. 4 (Apr. 1994), pp. 427–433. ISSN: 0020-9996. DOI: 10.1097/00004424-199404000-00008.

[87] M. P. Schilling, T. Scherr, F. R. Munke, et al. *Automated Annotator Variability Inspection for Biomedical Image Segmentation.* In: *IEEE Access*, Vol. 10 (2022), pp. 2753–2765. ISSN: 2169-3536. DOI: 10.1109/access.2022.3140378.

[88] L. Wiggins, P. J. O'Toole, W. J. Brackenbury, and J. Wilson. *Exploring the Impact of Variability in Cell Segmentation and Tracking Approaches.* In: *Microscopy Research and Technique*, Vol. 88, No. 3 (Nov. 2024), pp. 716–731. ISSN: 1097-0029. DOI: 10.1002/jemt.24715.

[89] C. Cortes and V. Vapnik. *Support-vector networks.* In: *Machine Learning*, Vol. 20, No. 3 (Sept. 1995), pp. 273–297. ISSN: 1573-0565. DOI: 10.1007/bf00994018.

[90] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez. *A comprehensive survey on support vector machine classification: Applications, challenges and trends.* In: *Neurocomputing*, Vol. 408 (Sept. 2020), pp. 189–215. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2019.10.118.

[91] L. Breiman. *Random Forests.* In: *Machine Learning*, Vol. 45, No. 1 (2001), pp. 5–32. ISSN: 0885-6125. DOI: 10.1023/a:1010933404324.

[92] T. Chen and C. Guestrin. *XGBoost: A Scalable Tree Boosting System.* In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[93] A. Krizhevsky, I. Sutskever, and G. E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks.* In: *Advances in Neural Information Processing Systems 25.* Curran Associates, Inc., 2012, pp. 1097–1105.

[94] A. G. Howard, M. Zhu, B. Chen, et al. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.* In: *arXiv* (2017). DOI: 10.48550/ARXIV.1704.04861.

[95] K. He, X. Zhang, S. Ren, and J. Sun. *Deep Residual Learning for Image Recognition.* In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778. DOI: `10.1109/cvpr.2016.90`.

[96] G. Huang, Z. Liu, L. VAN DER Maaten, and K. Q. Weinberger. *Densely Connected Convolutional Networks.* In: *arXiv* (2016). DOI: `10.48550/ARXIV.1608.06993`.

[97] C. Szegedy, W. Liu, Y. Jia, et al. *Going Deeper with Convolutions.* In: *arXiv* (2014). DOI: `10.48550/ARXIV.1409.4842`.

[98] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. *Aggregated Residual Transformations for Deep Neural Networks.* In: *arXiv* (2016). DOI: `10.48550/ARXIV.1611.05431`.

[99] K. Hara, H. Kataoka, and Y. Satoh. *Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition.* In: *arXiv* (2017). DOI: `10.48550/ARXIV.1708.07632`.

[100] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.* In: *arXiv* (2020). DOI: `10.48550/ARXIV.2010.11929`.

[101] Z. Liu, Y. Lin, Y. Cao, et al. *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows.* In: *arXiv* (2021). DOI: `10.48550/ARXIV.2103.14030`.

[102] H. Fan, B. Xiong, K. Mangalam, et al. *Multiscale Vision Transformers.* In: *arXiv* (2021). DOI: `10.48550/ARXIV.2104.11227`.

[103] S. Hochreiter and J. Schmidhuber. *Long Short-Term Memory.* In: *Neural Computation*, Vol. 9, No. 8 (Nov. 1997), pp. 1735–1780. ISSN: 1530-888X. DOI: `10.1162/neco.1997.9.8.1735`.

[104] J. Y.-H. Ng, M. Hausknecht, S. Vijayanarasimhan, et al. *Beyond short snippets: Deep networks for video classification.* In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE, June 2015, pp. 4694–4702. DOI: `10.1109/cvpr.2015.7299101`.

[105] N. Otsu. *A Threshold Selection Method.* In: *IEEE Transactions on Systems* (1979). DOI: `10.1109/tsmc.1979.4310076`.

[106] P.-S. LIAO, T.-S. Chen, and P.-C. Chung. *A Fast Algorithm for Multilevel Thresholding.* In: *Journal of Information Science and Engineering* (2001). DOI: `10.6688/jise.2001.17.5.1`.

[107] O. Ronneberger, P. Fischer, and T. Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* In: *arXiv* (2015). DOI: `10.48550/arxiv.1505.04597`.

[108] T. Scherr, K. Streule, A. Bartschat, et al. *BeadNet: deep learning-based bead detection and counting in low-resolution microscopy images.* In: *Bioinformatics*, Vol. 36, No. 17 (June 2020), pp. 4668–4670. ISSN: 1367-4811. DOI: `10.1093/bioinformatics/btaa594`.

[109] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu. *Cellpose: a generalist algorithm for cellular segmentation.* In: *Nature Methods*, Vol. 18, No. 1 (Dec. 2020), pp. 100–106. ISSN: 1548-7105. DOI: `10.1038/s41592-020-01018-x`.

[110] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.* In: *arXiv* (2016). DOI: `10.48550/arxiv.1606.06650`.

[111] E. Xie, W. Wang, Z. Yu, et al. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers.* In: *arXiv* (2021). DOI: `10.48550/arxiv.2105.15203`.

[112] A. Kirillov, E. Mintun, N. Ravi, et al. *Segment Anything.* In: *arXiv* (2023). DOI: `10.48550/ARXIV.2304.02643`.

[113] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim. *elastix: A Toolbox for Intensity-Based Medical Image Registration.* In: *IEEE Transactions on Medical Imaging*, Vol. 29, No. 1 (Jan. 2010), pp. 196–205. ISSN: 1558-254X. DOI: `10.1109/tmi.2009.2035616`.

[114] J. J. VAN Griethuysen, A. Fedorov, C. Parmar, et al. *Computational radiomics system to decode the radiographic phenotype.* In: *Cancer Research*, Vol. 77, No. 21 (Oct. 2017). DOI: `10.1158/0008-5472.can-17-0339`.

[115] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, et al. *scikit-image: image processing in Python.* In: *PeerJ*, Vol. 2 (2014), e453.

[116] M. Caron, H. Touvron, I. Misra, et al. *Emerging Properties in Self-Supervised Vision Transformers.* In: *arXiv* (2021). DOI: `10.48550/ARXIV.2104.14294`.

[117] L. McInnes, J. Healy, N. Saul, and L. Großberger. *UMAP: Uniform Manifold Approximation and Projection.* In: *Journal of Open Source Software*, Vol. 3, No. 29 (2018), p. 861. DOI: `10.21105/joss.00861`.

[118] H. Huang, Y. Wang, C. Rudin, and E. P. Browne. *Towards a comprehensive evaluation of dimension reduction methods for transcriptomic data visualization.* In: *Communications Biology*, Vol. 5, No. 1 (July 2022). ISSN: 2399-3642. DOI: `10.1038/s42003-022-03628-x`.

[119] M. A. Morid, A. Borjali, and G. Del Fiol. *A scoping review of transfer learning research on medical image analysis using ImageNet.* In: *Computers in Biology and Medicine*, Vol. 128 (Jan. 2021), p. 104115. ISSN: 0010-4825. DOI: `10.1016/j.compbiomed.2020.104115`.

[120] R. Azad, M. Heidary, K. Yilmaz, et al. *Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook.* In: *arXiv* (2023). DOI: `10.48550/ARXIV.2312.05391`.

[121] S. A. Taghanaki, Y. Zheng, S. K. Zhou, et al. *Combo Loss: Handling Input and Output Imbalance in Multi-Organ Segmentation.* In: *arXiv* (2018). DOI: `10.48550/ARXIV.1805.02798`.

[122] C. Shorten and T. M. Khoshgoftaar. *A survey on Image Data Augmentation for Deep Learning*. In: *Journal of Big Data*, Vol. 6, No. 1 (July 2019). ISSN: 2196-1115. DOI: `10.1186/s40537-019-0197-0`.

[123] H. Robbins and S. Monro. *A Stochastic Approximation Method.* In: *The Annals of Mathematical Statistics*, Vol. 22, No. 3 (Sept. 1951), pp. 400–407. ISSN: 0003-4851. DOI: `10.1214/aoms/1177729586`.

[124] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization.* In: *arXiv* (2014). DOI: `10.48550/ARXIV.1412.6980`.

[125] I. Loshchilov and F. Hutter. *Decoupled Weight Decay Regularization.* In: *arXiv* (2017). DOI: `10.48550/ARXIV.1711.05101`.

[126] A. Krogh and J. A. Hertz. *A simple weight decay can improve generalization.* In: *Proceedings of the 5th International Conference on Neural Information Processing Systems.* NIPS'91. Denver, Colorado: Morgan Kaufmann Publishers Inc., 1991, pp. 950–957. ISBN: 1558602224.

[127] S. Bos and E. Chug. *Using weight decay to optimize the generalization ability of a perceptron.* In: *Proceedings of International Conference on Neural Networks (ICNN'96).* Vol. 1. ICNN-96. IEEE, pp. 241–246. DOI: `10.1109/icnn.1996.548898`.

[128] R. R. Selvaraju, M. Cogswell, A. Das, et al. *Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization.* In: *2017 IEEE International Conference on Computer Vision (ICCV).* IEEE, Oct. 2017, pp. 618–626. DOI: `10.1109/iccv.2017.74`.

[129] J. Schindelin, I. Arganda-Carreras, E. Frise, et al. *Fiji: an open-source platform for biological-image analysis.* In: *Nature Methods*, Vol. 9, No. 7 (June 2012), pp. 676–682. ISSN: 1548-7105. DOI: `10.1038/nmeth.2019`.

[130] C. McQuin, A. Goodman, V. Chernyshev, et al. *CellProfiler 3.0: Next-generation image processing for biology.* In: *PLoS Biology*, Vol. 16, No. 7 (2018), e2005970. ISSN: 1544-9173. DOI: `10.1371/journal.pbio.2005970`.

[131] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, et al. *CellProfiler: image analysis software for identifying and quantifying cell phenotypes.* In: *Genome Biology*, Vol. 7, No. 10 (2006), R100–R100. ISSN: 1465-6906. DOI: `10.1186/gb-2006-7-10-r100`.

[132] I. Lukonin, D. Serra, L. Challet Meylan, et al. *Phenotypic landscape of intestinal organoid regeneration.* In: *Nature*, Vol. 586, No. 7828 (Oct. 2020), pp. 275–280. ISSN: 1476-4687. DOI: `10.1038/s41586-020-2776-9`.

[133] G. Gomez-Giro, J. Arias-Fuenzalida, J. Jarazo, et al. *Synapse alterations precede neuronal damage and storage pathology in a human cerebral organoid model of CLN3-juvenile neuronal ceroid lipofuscinosis.* In: *Acta Neuropathologica Communications*, Vol. 7, No. 1 (Dec. 2019). ISSN: 2051-5960. DOI: `10.1186/s40478-019-0871-7`.

[134] J. O. de Jong, C. Llapashtica, M. Genestine, et al. *Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder.* In: *Nature Communications*, Vol. 12, No. 1 (Sept. 2021). issn: 2041-1723. doi: `10.1038/s41467-021-24358-4`.

[135] Y. Huang, Z. Huang, Z. Tang, et al. *Research Progress, Challenges, and Breakthroughs of Organoids as Disease Models.* In: *Frontiers in Cell and Developmental Biology*, Vol. 9 (2021), p. 740574. issn: 2296-634X. doi: `10.3389/fcell.2021.740574`.

[136] A. Akshay, M. Katoch, M. Abedi, et al. *SpheroScan: A User-Friendly Deep Learning Tool for Spheroid Image Analysis.* In: *GigaScience*, Vol. 12 (Dec. 2022). issn: 2047-217X. doi: `10.1093/gigascience/giad082`.

[137] M. Streller, S. Michlikova, W. Ciecior, et al. *Image segmentation of treated and untreated tumor spheroids by fully convolutional networks.* In: *GigaScience*, Vol. 14 (2025). issn: 2047-217X. doi: `10.1093/gigascience/giaf027`.

[138] T. K. N. Ngo, S. J. Yang, B.-H. Mao, et al. *A deep learning-based pipeline for analyzing the influences of interfacial mechanochemical microenvironments on spheroid invasion using differential interference contrast microscopic images.* In: *Materials Today Bio*, Vol. 23 (Dec. 2023), p. 100820. issn: 2590-0064. doi: `10.1016/j.mtbio.2023.100820`.

[139] N. Hacıosmanoğlu, M. A. Güngen, E. G. Yilmaz, et al. *SpAi: A machine-learning supported experimental workflow for high-throughput spheroid production and analysis.* In: *Biosensors and Bioelectronics: X*, Vol. 23 (May 2025), p. 100588. issn: 2590-1370. doi: `10.1016/j.biosx.2025.100588`.

[140] M. Vitacolonna, R. Bruch, R. Schneider, et al. *A spheroid whole mount drug testing pipeline with machine-learning based image analysis identifies cell-type specific differences in drug efficacy on a single-cell level.* In: *BMC Cancer*, Vol. 24, No. 1 (Dec. 2024). issn: 1471-2407. doi: `10.1186/s12885-024-13329-9`.

[141] Z. Chen, N. Ma, X. Sun, et al. *Automated evaluation of tumor spheroid behavior in 3D culture using deep learning-based recognition.* In: *Biomaterials*, Vol. 272 (May 2021), p. 120770. issn: 0142-9612. doi: `10.1016/j.biomaterials.2021.120770`.

[142] M. P. Schilling, R. El Khaled El Faraj, J. E. Urrutia Gómez, et al. *Automated high-throughput image processing as part of the screening platform for personalized oncology.* In: *Scientific Reports*, Vol. 13, No. 1 (Mar. 2023). issn: 2045-2322. doi: `10.1038/s41598-023-32144-z`.

[143] J. Boyd, N. Lee, and J. Zoeller. *Automated High Content Imaging and Analysis of Spheroids Using Clearing and Deep Learning for Volumetric Quantification.* In: *Microscopy and Microanalysis*, Vol. 30 (July 2024). issn: 1435-8115. doi: `10.1093/mam/ozae044.168`.

[144] K. J. Trettner, J. Hsieh, W. Xiao, J. S. H. Lee, and A. M. Armani. *Nondestructive, quantitative viability analysis of 3D tissue cultures using machine learning image segmentation.* In: *APL Bioengineering*, Vol. 8, No. 1 (Mar. 2024). ISSN: 2473-2877. DOI: `10.1063/5.0189222`.

[145] W. H. Abuwatfa, W. G. Pitt, and G. A. Husseini. *Scaffold-based 3D cell culture models in cancer research.* In: *Journal of Biomedical Science*, Vol. 31, No. 1 (Jan. 2024). ISSN: 1423-0127. DOI: `10.1186/s12929-024-00994-y`.

[146] T. Kassis, V. Hernandez-Gordillo, R. Langer, and L. G. Griffith. *OrgaQuant: Human Intestinal Organoid Localization and Quantification Using Deep Convolutional Neural Networks.* In: *Scientific Reports*, Vol. 9, No. 1 (2019), p. 12479. DOI: `10.1038/s41598-019-48874-y`.

[147] B. Leng, H. Jiang, B. Wang, J. Wang, and G. Luo. *Deep-Orga: An improved deep learning-based lightweight model for intestinal organoid detection.* In: *Computers in Biology and Medicine*, Vol. 169 (Feb. 2024), p. 107847. ISSN: 0010-4825. DOI: `10.1016/j.compbiomed.2023.107847`.

[148] X. Bian, G. Li, C. Wang, et al. *A deep learning model for detection and tracking in high-throughput images of organoid.* In: *Computers in Biology and Medicine*, Vol. 134 (July 2021), p. 104490. ISSN: 0010-4825. DOI: `10.1016/j.compbiomed.2021.104490`.

[149] X. Xing, C. Tang, Y. Guo, N. Kurniawan, and G. Yang. *SegmentAnything helps microscopy images based automatic and quantitative organoid detection and analysis.* In: *arXiv* (2023). DOI: `10.48550/ARXIV.2309.04190`.

[150] M. A. Borten, S. S. Bajikar, N. Sasaki, H. Clevers, and K. A. Janes. *Automated brightfield morphometry of 3D organoid populations by OrganoSeg.* In: *Scientific Reports*, Vol. 8, No. 1 (2018), p. 5319. DOI: `10.1038/s41598-017-18815-8`.

[151] Z. Hasnain, A. K. Fraser, D. Georgess, et al. *OrgDyn: Feature and model based characterization of spatial and temporal organoid dynamics.* In: *Bioinformatics*, Vol. 36, No. 10 (2020), pp. 3292–3294. ISSN: 1367-4803. DOI: `10.1093/bioinformatics/btaa096`.

[152] A. Haja, J. M. Horcas-Nieto, B. M. Bakker, and L. Schomaker. *Towards automatization of organoid analysis: A deep learning approach to localize and quantify organoid images.* In: *Computer Methods and Programs in Biomedicine Update*, Vol. 3 (2023), p. 100101. ISSN: 2666-9900. DOI: `10.1016/j.cmpbup.2023.100101`.

[153] R. N. U. Kok, L. Hebert, G. Huelsz-Prince, et al. *OrganoidTracker: Efficient cell tracking using machine learning and manual error correction.* In: *PLoS ONE*, Vol. 15, No. 10 (2020), e0240802. DOI: `10.1371/journal.pone.0240802`.

[154] M. MacDonald, R. Fennel, A. Singanamalli, et al. *Improved automated segmentation of human kidney organoids using deep convolutional neural networks.* In: *Medical Imaging 2020: Image Processing*. SPIE, Mar. 2020, p. 118. DOI: `10.1117/12.2549830`.

[155] K. Park, J. Y. Lee, S. Y. Lee, et al. *Deep learning predicts the differentiation of kidney organoids derived from human induced pluripotent stem cells.* In: *Kidney Research and Clinical Practice*, Vol. 42, No. 1 (Jan. 2023), pp. 75–85. ISSN: 2211-9140. DOI: `10.23876/j.krcp.22.017`.

[156] J. Presberger, R. Keshara, D. Stein, et al. *Correlation Clustering of Organoid Images.* In: *arXiv* (2024). DOI: `10.48550/ARXIV.2403.13376`.

[157] P. Mergenthaler, S. Hariharan, J. M. Pemberton, et al. *Rapid 3D phenotypic analysis of neurons and organoids using data-driven cell segmentation-free machine learning.* In: *PLoS Computational Biology*, Vol. 17, No. 2 (2021), e1008630. ISSN: 1553-734X. DOI: `10.1371/journal.pcbi.1008630`.

[158] M. Salvi, U. Morbiducci, F. Amadeo, et al. *Automated Segmentation of Fluorescence Microscopy Images for 3D Cell Detection in human-derived Cardiospheres.* In: *Scientific Reports*, Vol. 9, No. 1 (2019), p. 6644. DOI: `10.1038/s41598-019-43137-2`.

[159] J. M. Matthews, B. Schuster, S. S. Kashaf, et al. *OrganoID: A versatile deep learning platform for tracking and analysis of single-organoid dynamics.* In: *PLOS Computational Biology*, Vol. 18, No. 11 (2022), e1010584. ISSN: 1553-734X. DOI: `10.1371/journal.pcbi.1010584`.

[160] L. Abdul, J. Xu, A. Sotra, et al. *D-CryptO: deep learning-based analysis of colon organoid morphology from brightfield images.* In: *Lab on a Chip*, Vol. 22, No. 21 (2022), pp. 4118–4128. ISSN: 1473-0189. DOI: `10.1039/d2lc00596d`.

[161] P. Mukashyaka, P. Kumar, D. J. Mellert, et al. *Cellos : High-throughput deconvolution of 3D organoid dynamics at cellular resolution for cancer pharmacology.* In: *bioRxiv* (2023). DOI: `10.1101/2023.03.03.531019`.

[162] J. D. Buehler, C. E. Bird, M. R. Savani, et al. *Semi-Automated Computational Assessment of Cancer Organoid Viability Using Rapid Live-Cell Microscopy.* In: *Cancer Informatics*, Vol. 21 (2022), p. 11769351221100754. ISSN: 1176-9351. DOI: `10.1177/11769351221100754`.

[163] Y. Tanaka, B. Cakir, Y. Xiang, G. J. Sullivan, and I.-H. Park. *Synthetic Analyses of Single-Cell Transcriptomes from Multiple Brain Organoids and Fetal Brain.* In: *Cell Reports*, Vol. 30, No. 6 (Feb. 2020), 1682–1689.e3. ISSN: 2211-1247. DOI: `10.1016/j.celrep.2020.01.038`.

[164] C. Luo, M. A. Lancaster, R. Castanon, et al. *Cerebral Organoids Recapitulate Epigenomic Signatures of the Human Fetal Brain.* In: *Cell Reports*, Vol. 17, No. 12 (Dec. 2016), pp. 3369–3384. ISSN: 2211-1247. DOI: `10.1016/j.celrep.2016.12.001`.

[165] S. Velasco, A. J. Kedaigle, S. K. Simmons, et al. *Individual brain organoids reproducibly form cell diversity of the human cerebral cortex.* In: *Nature*, Vol. 570, No. 7762 (June 2019), pp. 523–527. ISSN: 1476-4687. DOI: `10.1038/s41586-019-1289-x`.

[166] N. Gritti, J. L. Lim, K. Anlaş, et al. *MOrgAna: accessible quantitative analysis of organoids with machine learning.* In: *Development*, Vol. 148, No. 18 (2021), dev199611. ISSN: 0950-1991. DOI: `10.1242/dev.199611`.

[167] C. Brémond Martin, C. Simon Chane, C. Clouchoux, and A. Histace. *Mu-Net a Light Architecture for Small Dataset Segmentation of Brain Organoid Bright-Field Images*. In: *Biomedicines*, Vol. 11, No. 10 (Sept. 2023), p. 2687. ISSN: 2227-9059. DOI: `10.3390/biomedicines11102687`.

[168] C. Brémond-Martin, C. Simon-Chane, C. Clouchoux, and A. Histace. *TDA-Clustering Strategies for the Characterization of Brain Organoids*. In: *Ethical and Philosophical Issues in Medical Imaging, Multimodal Learning and Fusion Across Scales for Clinical Decision Support, and Topological Data Analysis for Biomedical Imaging*. Springer Nature Switzerland, 2022, pp. 113–122. ISBN: 9783031232237. DOI: `10.1007/978-3-031-23223-7_10`.

[169] J. Shi. *Automatic Quantitative Analysis of Brain Organoids via Deep Learning*. In: *arXiv* (2022). DOI: `10.48550/ARXIV.2211.00750`.

[170] B. Kahveci, E. Polatli, A. E. Evranos, et al. *BrAIn: A comprehensive artificial intelligence-based morphology analysis system for brain organoids and neuroscience*. In: *bioRxiv* (Feb. 2025). DOI: `10.1101/2025.02.19.638973`.

[171] K. Garg, A. Dev, P. Bansal, and H. Mittal. *An Efficient Deep Learning Model for Embryo Classification*. In: *2024 14th International Conference on Cloud Computing, Data Science and Engineering (Confluence)*. IEEE, Jan. 2024, pp. 358–363. DOI: `10.1109/confluence60223.2024.10463427`.

[172] D. Čapek, M. Safroshkin, H. Morales-Navarrete, et al. *EmbryoNet: using deep learning to link embryonic phenotypes to signaling pathways*. In: *Nature Methods*, Vol. 20, No. 6 (May 2023), pp. 815–823. ISSN: 1548-7105. DOI: `10.1038/s41592-023-01873-4`.

[173] N. Toulany, H. Morales-Navarrete, D. Čapek, et al. *Uncovering developmental time and tempo using deep learning*. In: *Nature Methods*, Vol. 20, No. 12 (Nov. 2023), pp. 2000–2010. ISSN: 1548-7105. DOI: `10.1038/s41592-023-02083-8`.

[174] A. J. R. Pond, S. Hwang, B. Verd, and B. Steventon. *A deep learning approach for staging embryonic tissue isolates with small data*. In: *PLOS ONE*, Vol. 16, No. 1 (Jan. 2021), e0244151. ISSN: 1932-6203. DOI: `10.1371/journal.pone.0244151`.

[175] T. Naert, Ö. Çiçek, P. Ogar, et al. *Deep learning is widely applicable to phenotyping embryonic development and disease*. In: *Development*, Vol. 148, No. 21 (Nov. 2021). ISSN: 1477-9129. DOI: `10.1242/dev.199664`.

[176] Y. Tokuoka, T. G. Yamada, D. Mashiko, et al. *3D convolutional neural networks-based segmentation to acquire quantitative criteria of the nucleus during mouse embryogenesis*. In: *npj Systems Biology and Applications*, Vol. 6, No. 1 (Oct. 2020). ISSN: 2056-7189. DOI: `10.1038/s41540-020-00152-8`.

[177] Z. Fan, Z. Li, Y. Jin, et al. *deepBlastoid: A Deep Learning-Based High-Throughput Classifier for Human Blastoids Using Brightfield Images with Confidence Assessment*. In: *bioRxiv* (Dec. 2024). DOI: `10.1101/2024.12.05.627041`.

[178] L. Deininger, S. Jung-Klawitter, R. Mikut, et al. *An AI-based segmentation and analysis pipeline for high-field MR monitoring of cerebral organoids.* In: *Scientific Reports*, Vol. 13, No. 1 (Dec. 2023). ISSN: 2045-2322. DOI: `10.1038/s41598-023-48343-7`.

[179] B. Vachha and S. Y. Huang. *MRI with ultrahigh field strength and high-performance gradients: challenges and opportunities for clinical neuroimaging at 7 T and beyond.* In: *European Radiology Experimental*, Vol. 5, No. 1 (Aug. 2021). ISSN: 2509-9280. DOI: `10.1186/s41747-021-00216-2`.

[180] A. S. Perera Molligoda Arachchige. *Neuroimaging with PET/MR: moving beyond 3 T in preclinical systems, when for clinical practice?* In: *Clinical and Translational Imaging*, Vol. 11, No. 4 (May 2023), pp. 315–319. ISSN: 2281-7565. DOI: `10.1007/s40336-023-00572-6`.

[181] milesial. *Pytorch-UNet: U-Net: Semantic segmentation with PyTorch.* `https://github.com/milesial/Pytorch-UNet`. 2020.

[182] A. Wolny, L. Cerrone, A. Vijayan, et al. *Accurate and versatile 3D segmentation of plant tissues at cellular resolution.* In: *eLife*, Vol. 9 (July 2020), e57613. ISSN: 2050-084X. DOI: `10.7554/eLife.57613`.

[183] O. Köpüklü, N. Kose, A. Gunduz, and G. Rigoll. *Resource efficient 3D convolutional neural networks.* In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).* IEEE. 2019, pp. 1910–1919.

[184] MMSegmentation Contributors. *MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark.* `https://github.com/open-mmlab/mmsegmentation`. 2020.

[185] A. Paszke, S. Gross, F. Massa, et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library.* In: *Advances in Neural Information Processing Systems.* Vol. 32. Curran Associates, Inc., 2019.

[186] A. d. Brebisson and G. Montana. *Deep Neural Networks for Anatomical Brain Segmentation.* In: *arXiv* (2015). DOI: `10.48550/arxiv.1502.02445`.

[187] R. Mehta, A. Majumdar, and J. Sivaswamy. *BrainSegNet: a convolutional neural network architecture for automated segmentation of human brain structures.* In: *Journal of Medical Imaging*, Vol. 4, No. 2 (2017), pp. 024003–024003. ISSN: 2329-4302. DOI: `10.1117/1.jmi.4.2.024003`.

[188] C. Wachinger, M. Reuter, and T. Klein. *DeepNAT: Deep convolutional neural network for segmenting neuroanatomy.* In: *NeuroImage*, Vol. 170 (2018), pp. 434–445. ISSN: 1053-8119. DOI: `10.1016/j.neuroimage.2017.02.035`.

[189] C.-C. C. Chen, J.-W. Chai, H.-C. Chen, et al. *An Iterative Mixed Pixel Classification for Brain Tissues and White Matter Hyperintensity in Magnetic Resonance Imaging.* In: *IEEE Access*, Vol. 7 (2019), pp. 124674–124687. ISSN: 2169-3536. DOI: `10.1109/access.2019.2931761`.

[190] S. Bao and A. C. S. Chung. *Multi-scale structured CNN with label consistency for brain MR image segmentation.* In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, Vol. 6, No. 1 (2018), pp. 113–117. ISSN: 2168-1163. DOI: 10.1080/21681163.2016.1182072.

[191] P. P. Garcez, E. C. Loiola, R. M. d. Costa, et al. *Zika virus impairs growth in human neurospheres and brain organoids.* In: *Science*, Vol. 352, No. 6287 (2016), pp. 816–818. ISSN: 0036-8075. DOI: 10.1126/science.aaf6116.

[192] N. Sun, X. Meng, Y. Liu, et al. *Applications of brain organoids in neurodevelopment and neurological diseases.* In: *Journal of Biomedical Science*, Vol. 28, No. 1 (2021), p. 30. ISSN: 1021-7770. DOI: 10.1186/s12929-021-00728-4.

[193] M. Scheibel, C. Nikulka, A. Dick, et al. *Structural Integrity and Clinical Function of the Subscapularis Musculotendinous Unit after Arthroscopic and Open Shoulder Stabilization.* In: *The American Journal of Sports Medicine*, Vol. 35, No. 7 (2007), pp. 1153–1161. ISSN: 0363-5465. DOI: 10.1177/0363546507299446.

[194] M. Scheibel, A. Tsynman, P. Magosch, R. J. Schroeder, and P. Habermeyer. *Postoperative Subscapularis Muscle Insufficiency after Primary and Revision Open Shoulder Stabilization.* In: *The American Journal of Sports Medicine*, Vol. 34, No. 10 (2006), pp. 1586–1593. ISSN: 0363-5465. DOI: 10.1177/0363546506288852.

[195] T. Eelbode, J. Bertels, M. Berman, et al. *Optimization for Medical Image Segmentation: Theory and Practice When Evaluating With Dice Score or Jaccard Index.* In: *IEEE Transactions on Medical Imaging*, Vol. 39, No. 11 (2020), pp. 3679–3690. ISSN: 0278-0062. DOI: 10.1109/tmi.2020.3002417.

[196] *ISLES Challenge 2017.* http://www.isles-challenge.org/ISLES2017/. Accessed: 2023-04-05.

[197] *ISLES Challenge 2018.* http://www.isles-challenge.org/ISLES2018/. Accessed: 2023-04-05.

[198] J. Schröter, L. Deininger, B. Lupse, et al. *A large and diverse brain organoid dataset of 1,400 cross-laboratory images of 64 trackable brain organoids.* In: *Scientific Data*, Vol. 11, No. 1 (May 2024). ISSN: 2052-4463. DOI: 10.1038/s41597-024-03330-z.

[199] J. Schröter, H. Syring, G. Göhring, et al. *Generation of an induced pluripotent stem cell line (DHMCi008-A) from an individual with TUBA1A tubulinopathy.* In: *Stem Cell Research*, Vol. 62 (July 2022), p. 102818. ISSN: 1873-5061. DOI: 10.1016/j.scr.2022.102818.

[200] J. Schröter, H. Syring, G. Göhring, et al. *Generation of an induced pluripotent stem cell line (DHMCi009-A) from an individual with TUBB2A tubulinopathy.* In: *Stem Cell Research*, Vol. 64 (Oct. 2022), p. 102879. ISSN: 1873-5061. DOI: 10.1016/j.scr.2022.102879.

[201] K. Fei, J. Zhang, J. Yuan, and P. Xiao. *Present Application and Perspectives of Organoid Imaging Technology.* In: *Bioengineering*, Vol. 9, No. 3 (Mar. 2022), p. 121. ISSN: 2306-5354. DOI: 10.3390/bioengineering9030121.

[202] A. P.-T. Jost and J. C. Waters. *Designing a rigorous microscopy experiment: Validating methods and avoiding bias.* In: *Journal of Cell Biology*, Vol. 218, No. 5 (Mar. 2019), pp. 1452–1466. ISSN: 1540-8140. DOI: `10.1083/jcb.201812109`.

[203] L. Alzubaidi, J. Zhang, A. J. Humaidi, et al. *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.* In: *Journal of Big Data*, Vol. 8, No. 1 (Mar. 2021). ISSN: 2196-1115. DOI: `10.1186/s40537-021-00444-8`.

[204] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger. *Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control.* In: *NeuroImage*, Vol. 195 (July 2019), pp. 11–22. ISSN: 1053-8119. DOI: `10.1016/j.neuroimage.2019.03.042`.

[205] S. Handcock, K. Richards, T. J. Karle, et al. *Three-Dimensional Morphological Characterisation of Human Cortical Organoids Using a Customised Image Analysis Workflow.* In: *Organoids*, Vol. 4, No. 1 (Jan. 2025), p. 1. ISSN: 2674-1172. DOI: `10.3390/organoids4010001`.

[206] D. Hannuschke. *Master thesis: Single cell RNA sequencing analysis of patient derived cerebral organoids.* In: *Universität Heidelberg, Faculty of Engineering Sciences, Master Program Molecular Biotechnology* (Sept. 2023).

[207] M. D. Luecken and F. J. Theis. *Current best practices in single-cell RNA-seq analysis: a tutorial.* In: *Molecular Systems Biology*, Vol. 15, No. 6 (June 2019). ISSN: 1744-4292. DOI: `10.15252/msb.20188746`.

[208] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. *Fast unfolding of communities in large networks.* In: *arXiv* (2008). DOI: `10.48550/ARXIV.0803.0476`.

[209] Z. He and B. Treutlein. *scRNAseq analysis vignette.* `https://github.com/quadbio/scRNAseq_analysis_vignette/blob/master/Tutorial.md`. Accessed: 2025-01-24.

[210] D. Kachkin, N. Zagha, T. Boerstler, et al. *Deciphering Brain Organoids Heterogeneity by Identifying Key Quality Determinants.* In: *bioRxiv* (Jan. 2025). DOI: `10.1101/2025.01.13.632763`.

[211] G. X. Y. Zheng, J. M. Terry, P. Belgrader, et al. *Massively parallel digital transcriptional profiling of single cells.* In: *Nature Communications*, Vol. 8, No. 1 (Jan. 2017). ISSN: 2041-1723. DOI: `10.1038/ncomms14049`.

[212] C. Wolfe, Y. Feng, D. Chen, et al. *GeoTyper: Automated Pipeline from Raw scRNA-Seq Data to Cell Type Identification.* In: *arXiv* (2022). DOI: `10.48550/ARXIV.2205.01187`.

[213] Z. He, A. Maynard, A. Jain, et al. *Lineage recording in human cerebral organoids.* In: *Nature Methods*, Vol. 19, No. 1 (Dec. 2021), pp. 90–99. ISSN: 1548-7105. DOI: `10.1038/s41592-021-01344-8`.

[214] G. Lozachmeur, A. Bramoulle, A. Aubert, et al. *Three-dimensional molecular cartography of human cerebral organoids revealed by double-barcoded spatial transcriptomics.* In: *Cell Reports Methods*, Vol. 3, No. 9 (Sept. 2023), p. 100573. ISSN: 2667-2375. DOI: `10.1016/j.crmeth.2023.100573`.

[215] D. Vonficht, L. Jopp-Saile, S. Yousefian, et al. *Ultra-high scale cytometry-based cellular interaction mapping*. Accepted at Nature Methods. 2025.

[216] E. V. Rothenberg. *Transcriptional Control of Early T and B Cell Developmental Choices*. In: *Annual Review of Immunology*, Vol. 32, No. 1 (Mar. 2014), pp. 283–321. ISSN: 1545-3278. DOI: `10.1146/annurev-immunol-032712-100024`.

[217] S. Marjan, T. Yangzom, C. K. Kristiansen, et al. *Application of Flow Cytometric Analysis for Measuring Multiple Mitochondrial Parameters in 3D Brain Organoids*. In: *Journal of Visualized Experiments*, No. 198 (Aug. 2023). ISSN: 1940-087X. DOI: `10.3791/65621`.

[218] R. A. Thomas, J. Sirois, S. Li, et al. *CelltypeR: A flow cytometry pipeline to characterize single cells from brain organoids*. In: *iScience*, Vol. 27, No. 9 (Sept. 2024), p. 110613. ISSN: 2589-0042. DOI: `10.1016/j.isci.2024.110613`.

[219] N. F. Greenwald, G. Miller, E. Moen, et al. *Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning*. In: *Nature Biotechnology*, Vol. 40, No. 4 (Nov. 2021), pp. 555–565. ISSN: 1546-1696. DOI: `10.1038/s41587-021-01094-0`.

[220] G. W. Zack, W. E. Rogers, and S. A. Latt. *Automatic measurement of sister chromatid exchange frequency*. In: *Journal of Histochemistry and Cytochemistry*, Vol. 25, No. 7 (July 1977), pp. 741–753. ISSN: 1551-5044. DOI: `10.1177/25.7.70454`.

[221] E. Rodríguez-Correa, F. Grünschläger, T. Nizharadze, et al. *A kinetics-based model of hematopoiesis reveals extrinsic regulation of skewed lineage output from stem cells*. In: *bioRxiv* (Feb. 2025). DOI: `10.1101/2025.02.04.636388`.

[222] T. Peng, K. Thorn, T. Schroeder, et al. *A BaSiC tool for background and shading correction of optical microscopy images*. In: *Nature Communications*, Vol. 8, No. 1 (June 2017). ISSN: 2041-1723. DOI: `10.1038/ncomms14836`.

[223] F. C. A. Groen, I. T. Young, and G. Ligthart. *A comparison of different focus functions for use in autofocus algorithms*. In: *Cytometry*, Vol. 6, No. 2 (Mar. 1985), pp. 81–91. ISSN: 1097-0320. DOI: `10.1002/cyto.990060202`.

[224] J. Moss. *pyslingshot: A Python implementation of Slingshot trajectory inference*. `https://github.com/mossjacob/pyslingshot`. Accessed: 2025-05-24. 2023.

[225] K. Street, D. Risso, R. B. Fletcher, et al. *Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics*. In: *BMC Genomics*, Vol. 19, No. 1 (June 2018). ISSN: 1471-2164. DOI: `10.1186/s12864-018-4772-0`.

[226] M. A. Reiche, J. S. Aaron, U. Boehm, et al. *When light meets biology – how the specimen affects quantitative microscopy*. In: *Journal of Cell Science*, Vol. 135, No. 6 (Mar. 2022). ISSN: 1477-9137. DOI: `10.1242/jcs.259656`.

[227] G. Beretta, A. F. Mastorgio, L. Pedrali, S. Saponaro, and E. Sezenna. *The effects of electric, magnetic and electromagnetic fields on microorganisms in the perspective of bioremediation.* In: *Reviews in Environmental Science and Bio/Technology*, Vol. 18, No. 1 (Jan. 2019), pp. 29–75. ISSN: 1572-9826. DOI: `10.1007/s11157-018-09491-9`.

[228] A. Versace, T. K. Hitchens, C. T. Wallace, S. C. Watkins, and L. D'Aiuto. *11.7T Diffusion Magnetic Resonance Imaging and Tractography to Probe Human Brain Organoid Microstructure.* In: *Biological Psychiatry Global Open Science*, Vol. 4, No. 5 (2024), p. 100344. ISSN: 2667-1743. DOI: `https://doi.org/10.1016/j.bpsgos.2024.100344`.

[229] S. L. Giandomenico, M. Sutcliffe, and M. A. Lancaster. *Generation and long-term culture of advanced cerebral organoids for studying later stages of neural development.* In: *Nature Protocols*, Vol. 16, No. 2 (Dec. 2020), pp. 579–602. ISSN: 1750-2799. DOI: `10.1038/s41596-020-00433-w`.

[230] B. Phipson, P. X. Er, A. N. Combes, et al. *Evaluation of variability in human kidney organoids.* In: *Nature Methods*, Vol. 16, No. 1 (Dec. 2018), pp. 79–87. ISSN: 1548-7105. DOI: `10.1038/s41592-018-0253-2`.

[231] Z. Zhao, X. Chen, A. M. Dowbaj, et al. *Organoids.* In: *Nature Reviews Methods Primers*, Vol. 2, No. 1 (Dec. 2022). ISSN: 2662-8449. DOI: `10.1038/s43586-022-00174-y`.

[232] A. Louey, D. Hernández, A. Pébay, and M. Daniszewski. *Automation of Organoid Cultures: Current Protocols and Applications.* In: *SLAS Discovery*, Vol. 26, No. 9 (Oct. 2021), pp. 1138–1147. ISSN: 2472-5552. DOI: `10.1177/24725552211024547`.

[233] K. Sarieva, T. Kagermeier, V. Lysenkov, et al. *Comparing the impact of sample multiplexing approaches for single-cell RNA-sequencing on downstream analysis using cerebellar organoids.* In: *bioRxiv* (Aug. 2024). DOI: `10.1101/2024.08.23.609290`.

[234] S. Charles, E. Jackson-Holmes, G. Sun, et al. *Non-Invasive Quality Control of Organoid Cultures Using Mesofluidic CSTR Bioreactors and High-Content Imaging.* In: *Advanced Materials Technologies*, Vol. 10, No. 3 (Aug. 2024). ISSN: 2365-709X. DOI: `10.1002/admt.202400473`.

[235] K. Marzec-Schmidt, N. Ghosheh, S. R. Stahlschmidt, et al. *Artificial Intelligence Supports Automated Characterization of Differentiated Human Pluripotent Stem Cells.* In: *Stem Cells*, Vol. 41, No. 9 (June 2023), pp. 850–861. ISSN: 1549-4918. DOI: `10.1093/stmcls/sxad049`.

[236] G. N. Kanda, T. Tsuzuki, M. Terada, et al. *Robotic search for optimal cell culture in regenerative medicine.* In: *eLife*, Vol. 11 (June 2022), e77007. ISSN: 2050-084X. DOI: `10.7554/eLife.77007`.

[237] Z. Wu, J. Park, P. R. Steiner, B. Zhu, and J. X. J. Zhang. *Generative adversarial network model to classify human induced pluripotent stem cell-cardiomyocytes based on maturation level.* In: *Scientific Reports*, Vol. 14, No. 1 (Nov. 2024). ISSN: 2045-2322. DOI: `10.1038/s41598-024-77943-0`.

[238] W. Lei, C. Fuster-Barceló, G. Reder, A. Muñoz-Barrutia, and W. Ouyang. *BioImage.IO Chatbot: a community-driven AI assistant for integrative computational bioimaging*. In: *Nature Methods*, Vol. 21, No. 8 (Aug. 2024), pp. 1368–1370. ISSN: 1548-7105. DOI: `10.1038/s41592-024-02370-y`.