Modelling the Placenta from the Bottom Up: Digital Phenotyping of Human Placenta Histology

Claudia Vanea

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2024

Acknowledgements

Personal

I thank my amazing supervisors, Chris and Cecilia. Chris, your support, direction, attentiveness, (trolling), AI debates, and idea exploration have made this such a fun three years, and I'm really proud of everything we've accomplished. Cecilia, you are an inspirational leader and someone I have looked up to since I started working with you. I will take forward many of your lessons in my future endeavours. I have learnt so much under both of your supervision.

I thank Linda for her incredible mentorship and hard work. You taught me all about this crazy organ and helped to push the project towards clinical need. I might still be mixing up mature and immature intermediate villi without you.

I thank my parents for always supporting and believing in me, no matter which academic discipline or path I've tread. I thank the chillies, elf data scientists, and my old school friends for keeping me sane and entertained throughout all of this.

Tom. You have supported and encouraged me in everything, never doubting my abilities and always pushing me to be my best. You are my incredibly reliable, if at times impatient, tech support. I finally acknowledge that you came up with the very popular name for HAPPY, for which I stole all the credit. There's so much more I could say but I will save it for a another speech. Just, thank you.

Institutional

I thank the University of Oxford EPSRC Center for Doctoral Training in Health Data Science (HDS CDT) for their training and support, for encouraging a tight-knit community among cohorts, and for giving me the opportunity to teach new CDT students every year. I was supported by the EPSRC (EP/S02428X/1) and the computational aspects of my research by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z and the NIHR Oxford BRC.

List of Publications

The following is a list of articles and conference proceedings that were published as a result of the research carried out for this thesis.

Articles

- Vanea C, Džigurski J, Rukins V, Dodi O, Siigur S, Salumäe L, et al. Mapping cell-to-tissue graphs across human placenta histology whole slide images using deep learning with HAPPY. *Nat Commun.* 2024;15: 1–16.

Conference Proceedings

- Campbell J*, Vanea C*, Salumäe L, Meir K, Hochner-Celnikier D, Hochner H, et al. Enhancing Cross-Institute Generalisation of GNNs in Histopathology through Multiple Embedding Graph Augmentation (MEGA). preprint. (*equal contribution)

- Vanea C, Campbell J, Dodi O, Salumäe L, Meir K, Hochner-Celnikier D, et al. A New Graph Node Classification Benchmark: Learning Structure from Histology Cell Graphs. *New Frontiers in Graph Learning at NeurIPS 2022*; 2022. doi:10.48550/arXiv.2211.06292

- Yuan H*, Vanea C*, Lucivero F, Hallowell N. Training Ethically Responsible AI Researchers: a Case Study. *Navigating the Broader Impacts of AI Research at NeurIPS 2020.* 2020. doi:10.48550/arXiv.2011.11393. (*equal contribution)

Abstract

Accurate placenta histopathology assessment is essential for immediate and lifelong clinical management of mother and newborn health. However, the placenta's heterogeneity and high disease tolerance are challenging for robust, reproducible histological analysis and detection of pathological changes. Placenta histology slides contain upwards of a million cells and tens of thousands of micro-anatomical tissues. High-throughput, quantitative and objective metrics of placental biology are therefore valuable for placental investigations in clinical and research settings. Deep learning for digital pathology has the potential to provide these high-throughput metrics, but the placenta has received relatively little attention compared to other organ histology.

In this work, I present a three-stage hierarchical deep learning pipeline for analysing the placenta from the bottom up. The pipeline outputs rich and interpretable biological metrics following the organ's anatomical hierarchy. I start the pipeline at the cellular level, localising all nuclei across a slide and classifying them into one of 11 cell types. Using the nuclei coordinates and cell classes, I construct a whole slide cell graph that mirrors cellular community interaction within micro-anatomical tissue structures. I develop a scalable graph neural network which uses constituent cells to predict one of 9 tissue microstructures in the placenta parenchyma.

The whole slide cellular and tissue microstructure predictions match expectations from placental biology literature and closely replicate those from independent clinical experts. I use these whole slide metrics to quantify healthy variation in the term placenta and find that significant pathological changes are caused by common placental lesions. Finally, I develop a novel graph compression autoencoder with a custom compression algorithm for unsupervised region clustering. This work is a step towards automated, interpretable and high-throughput metrics for assessing placental health, with potential applications in both clinical and research settings.

Contents

List of Figures

List of Tables

Glossary

AI	Artificial intelligence including deep learning models.
H&E	Hematoxylin and eosin stain of histology slides.
MLP	Multi-layer perceptron model.
GPU, CPU	Graphics processing unit or central processing unit, hardware computing components.
CNN	Convolutional neural network.
GNN	Graph neural network.
GCN	Graph convolutional network.
LLM	Large language model.
WSI	Whole slide image, a digitised histology slide.
GB, TB	Gigabyte or terabyte.
patch, tile	Image portions of a whole slide image.
2D, 3D	Two- or three-dimensions.
MIL	Multiple instance learning.
FDA	Food and Drug Administration device approval.
CE-IVDR	European Union Conformité Européenne In Vitro Diagnostic device approval.
DNA, RNA	Deoxyribonucleic Acid and Ribonucleic Acid.
UoT	University of Tartu, one of our data collaborators and source of histology slides.
HMC	Hadassah Medical Center, one of our data collaborators and source of histology slides.
NUH	Northshore University HealthSystem, one of our data collaborators and source of histology slides.
µm	Micrometer, one-thousandth of a millimetre.
HDF5	Hierarchical Data Formats, an efficient on-disk storage format.

SQL	Structured query language, programming language for interacting with a relational database.
ReLU	Rectified linear unit, a non-linear activation function.
F1 Score	The harmonic mean of precision and recall scores.
RGB	Red, green, blue colour spectrum.
ROC AUC	Receiver operating characteristic area under curve, a performance measure.
Δ	Delta, used here to show the difference between two measures.
KNN	k-Nearest Neighbours.
UMAP	Uniform Manifold Approximation and Projection, a dimensionality reduction technique.
MAE, MSE	Mean absolute error or mean squared error.
HAPPY	Histology Analysis Pipeline in Python, the three-stage AI placenta analysis pipeline developed as part of this research.
SOP	Standard Operating Procedure.
PR-AUC	Precision recall area under curve, a performance measure.
GCAE	Graph compression autoencoder.
supernodes	Remaining nodes after graph compression.
childnodes	Nodes which were removed and compressed into supernodes with graph compression.
tanh	Hyperbolic tangent function used as an activation function.
GAT, GATv2	Graph Attention Network.
GIN	Graph Isomorphism Network.
PCC	Pearson Correlation Coefficient.

Know thyself.

1

Introduction

Contents

1.1 Motivation

Accurate placenta histopathology assessment is essential for immediate and life-long clinical management of mother and newborn health. Placenta pathology informs immediate clinical treatment, clinical intervention and management for long-term adverse maternal and child health outcomes, predicts recurrence risk in subsequent pregnancies and explains underlying causes of pregnancy loss [1–9]. From a research perspective, placental histopathology can determine new biomarkers of disease, such as the recently identified SARS-CoV-2 placentitis [10, 11], and improve our understanding of the biological mechanisms of healthy development and disease processes. There is increasing evidence for the impact of the placenta on fetal developmental origins of health and disease and changes to the maternal body with life-long health consequences [12–18].

The placenta is unique among organs for its transience, heterogeneity, and high disease tolerance, making it challenging for histological analysis. It is a rapidly changing organ with expected anatomical differences week by week where ‘normal’ can only be assessed in reference to gestational age [8]. It is heterogeneous both within the same sample and across samples [8, 19], with inherent genetic mosaicism not found anywhere else in the body [20]. The placenta exhibits a high reserve capacity, redundancy and adaptive capabilities with lesions present in placentas from uncomplicated pregnancies [8, 13, 21, 22], making it challenging to define a normal placenta. It has been said that “if you’ve seen one normal kidney, you’ve seen them all, but to see a normal placenta takes years of experience” [23].

The difficulty of placenta histology analysis is reflected in low inter-observer reliability among pathologists for placental features and pathologies [8, 19, 24–27]. This is especially true when comparing specialist perinatal pathologists to general surgical pathologists [28–30], with one study reporting that “there is no agreement between General Surgical Pathologists and Perinatal Pathologists when assessing placental pathology other than Acute Inflammation, and weak agreement even for Acute Inflammation” [29]. However, there are few perinatal pathologists relative to the number of submissions to placenta histology investigation [19, 31]. In the UK, the Royal College of Pathologists reports that there are fewer than 70 perinatal or paediatric pathologists [32] for around 600,000 annual births [33], where 20-40% of those are submitted for histological investigation [3, 34–36]. As a consequence, findings may be assessed by non-specialists and with 1 in 200 pregnancies resulting in stillbirth or neonatal death [37, 38], there can be long wait times to receive post-mortem results [37, 39]. As not all institutes have a pathology lab, if pathologists are unavailable, samples may be stored before being processed for pathology. However, the duration and type of storage can impact placental morphology and potentially influence diagnosis [22, 40, 41].

Challenges in the clinical setting are further compounded in the research setting. Despite its importance, the placenta has been a historically understudied organ [4, 13, 31, 42–45], suffering from a general lack of funding for reproductive biology research [45] and awareness in the public eye. The placenta's high heterogeneity, temporal variability and disease tolerance require large cohorts with rich whole slide phenotypes, but expert time to curate such data is limited. Placenta histopathology is traditionally studied using stereology, an unbiased sampling technique for generalising from manually identified features in randomly sampled sites [46–48]. However, the requirement for manual annotation can limit study sample sizes and either take up perinatal pathologists' time or have annotations generated by non-experts.

Artificial intelligence (AI) for digital pathology has the potential to provide high-throughput metrics to assist in clinical diagnosis and offer automated, comprehensive, and quantitative phenotypes for accelerating research [49]. However, the placenta has seen little attention in this space, with few methods explicitly developed for placenta histology [50–53]. A lack of open placenta histology repositories and datasets has further led to its exclusion from large multi-organ AI models. There is much that we still do not know about the placenta, and many highlight the need for more research [3, 8, 54, 55].

1.2 Contribution

In this thesis, I present a series of deep learning digital pathology models for comprehensive phenotyping of placenta histology. I approach this analysis 'from the bottom up', starting at the nucleus, building through cells and micro-anatomical tissue structures to the slide- and patient-level. The method outputs rich whole slide metrics at each scale, enforces a biological prior based on hierarchical placental anatomy, and is validated against explainable biology at every step. I use these whole slide metrics to quantify healthy variation in the term placenta and find that significant pathological changes are caused by common placental lesions. This research is a step towards outputting automated, interpretable and high-throughput metrics for assessing placenta health. It has the potential to be developed into a tool for assistive clinical and research histopathology, exploring the placenta's heterogeneity and disease tolerance and easing the burden on perinatal pathologists.

1.2.1 Thesis Summary

In **Chapter 2**, I provide the background for placental biology across relevant biological scales and describe current placental histopathology clinical practice. I present an overview of the major advancements in deep learning which led to the model architectures used in this work. Finally, I contextualise the research within the field of digital pathology.

In **Chapter 3**, I describe the start of the deep learning pipeline. After detailing our whole slide datasets and data collaborations, I present two deep learning models for efficient whole slide cellular phenotyping. The first model localises all nuclei across a slide, and the second classifies them into one of 11 placenta parenchyma cell types. I explore the effect of augmentations on model generalisability and transfer them to umbilical cord cells by fine-tuning with a small amount of data. This is the first method for automated whole slide cellular phenotyping in placenta histology and the first for classifying 11 different types of cell found in the placenta parenchyma.

In **Chapter 4**, I use the nuclei coordinates and cell classes to construct a whole slide cell graph that mirrors cellular community interaction within micro-anatomical tissue structures. I develop a scalable graph neural network which uses constituent cellular communities to predict one of 9 tissue microstructures in the placenta parenchyma. I show how this model generalises to slides from a held-out institute and present a comprehensive hyperparameter exploration. I compare the performance of seven scaleable state-of-the-art graph neural networks and describe a cell graph dataset release to the graph learning community for developing new scalable architectures. It is the first method for automated classification and localisation of tissue microstructures in placenta histology. It is one of few methods using graph neural networks for node classification on cell graphs built across entire whole slide images.

In **Chapter 5**, I use the prior models to generate aggregated whole slide cellular and tissue microstructure metrics for quantifying healthy variation in the term placenta. I validate the outputs against the agreement scores of four practising perinatal pathologists and expectations from biology and placental literature. The proportion of cells, tissue microstructures, and cellular composition of tissue microstructures matches expectations for healthy term placentas. This is the first automated method that combines cell and tissue microstructure outputs to quantify placenta histology and provides the first set of whole slide cell and tissue microstructure benchmarks for healthy term placentas.

In **Chapter 6**, I first explore the aggregated whole slide cell and tissue microstructure changes on slides with one or more of four clinically significant parenchymal lesions. I find that there are significant differences in the cells and tissue microstructures across slides with lesions and that these match the expected pathological changes caused by these lesions. I then present preliminary work on using a novel graph compression autoencoder with a custom compression algorithm for unsupervised region clustering. This model is a new architecture for feature reconstruction with a graph compression bottleneck and is the first application of graph autoencoders to histology cell graphs.

In **Chapter 7**, I describe future directions for this research. Immediate improvements can come from training with slides from earlier trimesters across all placenta slide types and lesions currently of clinical interest. The methodology explored in Chapter 6 can be extended with variants inspired by denoising diffusion models and graph transformers. With large enough placenta whole slide datasets, self-supervised phenotyping can be explored using methods like self distillation and diffusion autoencoders. In the long term, developing an automated placental histopathology software tool would provide the most real-world utility for assistive clinical placental pathology and enriching placenta histology research.

In **Chapter 8**, I present an ethics impact analysis for this research and general ethical considerations for biomedical collaborations. After describing the ethical approvals obtained for this research and other ethical considerations, I cover four potential fail states to account for in biomedical collaborations. These include the risk of overreliance on AI, a lack of clinical or AI domain knowledge during development and deployment, and the impact of using AI in clinical workflows on the patient.

It is so characteristic, that just when the mechanics of reproduction are so vastly improved, there are fewer and fewer people who know how the music should be played.

— Ludwig Wittgenstein

2

Background

Contents

2.1 Biology of the Placenta

2.1.1 Overview and Development

The placenta is the first organ that we form during pregnancy. It performs the roles of fetal organ systems as they develop, it protects the fetus from the maternal immune system, it regulates endocrine activity, and it facilitates the exchange of oxygen, nutrients and waste [13]. Unique cells of the placenta begin differentiating along with the embryo 5-8 days after fertilisation [56]. This ball of cells, together with the encased embryo, adheres to the top or side of the uterine wall at 5-6 days, where it remains as the placenta until it is delivered after birth. By day 13, the placenta begins to compartmentalise into its main structures. The embryonic surface becomes the chorionic plate, the maternal surface becomes the basal plate, and lacunae develop between the two, which fill with maternal blood. By day 35, the placenta takes on its characteristic disk morphology (Figure 2.1), no longer surrounding the developing fetus with a connection maintained by the umbilical cord. The tissues of the placenta continue to mature and change throughout pregnancy to meet the needs of the developing fetus.

The placenta has an unusual genetic profile, with inherent mosaicism [20], making it highly heterogeneous, and unique cell types that have incorporated ancient endogenous retroviral envelope genes for cell fusion [57–59]. The human placenta is most similar to those of great apes but is otherwise unique among mammals for its high degree of invasive placentation and internal branching villus structure. These qualities are thought to trade off longer gestational periods and fetal development for a higher risk of complications [13, 60].

2.1.2 Main Structures

At the start of the midtrimester, compartmentalised placental anatomy is established [61] and consists of four main structures: the umbilical cord, the basal plate (also called maternal surface), the chorionic plate (also called fetal surface), and the parenchyma (Figure 2.1). The umbilical cord connects the placenta to the fetus and contains two arteries and one vein, which handle all vascular exchange between the fetus and the placenta. The chorionic plate, the surface of the placenta facing the fetus, is the first branching point of these veins and arteries. From the chorionic plate, the chorionic villi grow into the maternal-blood-filled intervillous space that was previously the lacunae. On the other side of the placenta is the basal plate, which is the surface attached to decidualised maternal uterine tissue, anchoring the organ and providing access to maternal veins and arteries. The parenchyma, which is of focus in this research, is the region between the chorionic and basal

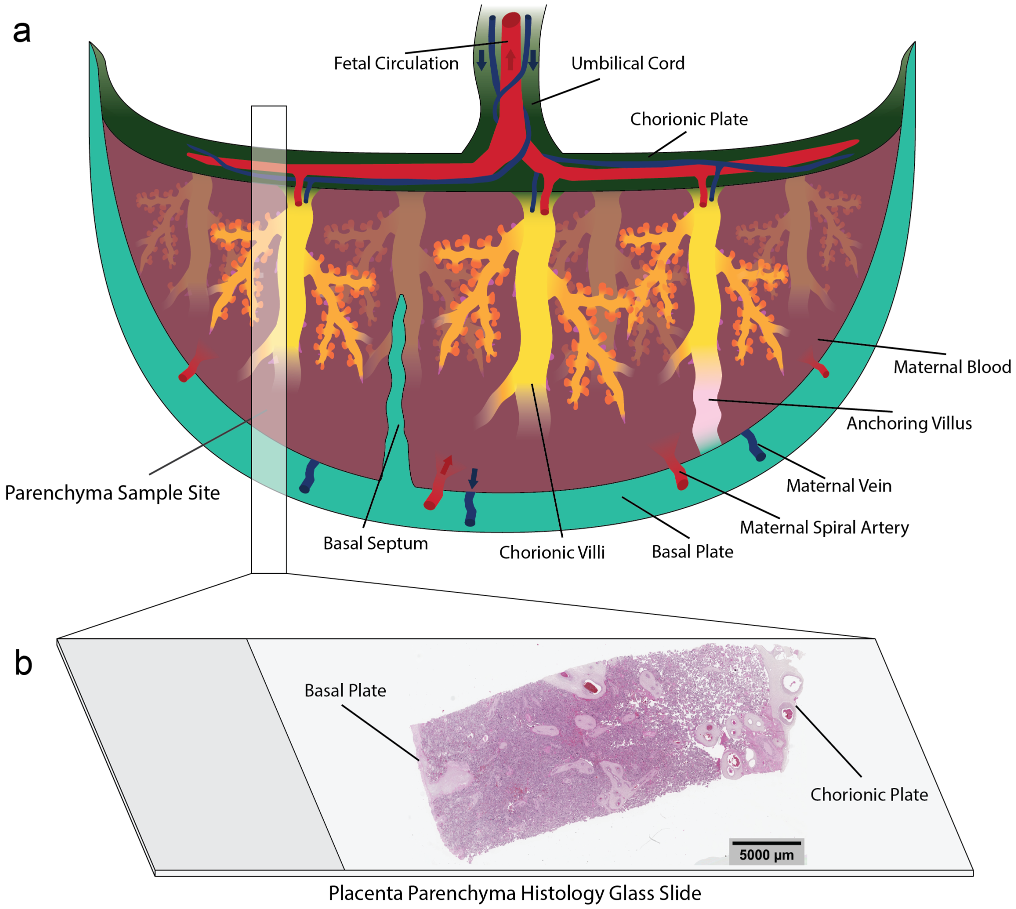


Figure 2.1: Schematic diagram of a healthy term placenta with a histology image on a glass slide. (a) Healthy term placenta schematic showing macrostructures including the basal plate, chorionic plate and branching chorionic villi (not drawn to scale). The fetal vasculature extends from the umbilical cord through the chorionic plate and into the chorionic villi for maternal/fetal diffusive exchange. The maternal spiral arteries supply the intervillous space with maternal blood, and the maternal veins remove deoxygenated blood and waste. An example parenchyma slide sample site is shown by the semi-transparent box. (b) The resulting histology section presented on a glass slide.

plates containing the chorionic villi and is the site of maternal/fetal diffusive vascular exchange (Figure 2.1).

2.1.3 Tissue Microstructures

The main structures of the placenta are composed of smaller tissue microstructures. Tissue microstructures, sometimes called functional tissue units, are micro-anatomical communities of cells within an organ which interact towards a higher-order function. They are the smallest organisation of tissue that performs a unique physiological function, replicated across the organ [62]. Examples throughout the body include the nephrons of the kidney, the intestinal villi of the small intestine, and the alveoli of the lungs. Categorisation and subcategorisation of tissue into microstructures is somewhat arbitrary as biology is nuanced and continuous. However, representing the biological organisation at a scale between the organ and its cells is useful for clinical diagnosis and biological understanding. Changes to the expected compositions of these tissue microstructures are indicators of a pathologic process.

Focusing on the placenta parenchyma as it is of primary interest in this work, tissue microstructures can be grouped into two main categories: the chorionic villi and the maternal/fetal surfaces [23, 61].

Chorionic Villi

The chorionic villi comprise the fetal vascular system within the placenta (Figure 2.2). Growing from the chorionic plate in a root or tree-like structure into the nutrient-rich maternal blood-filled space, they provide a high surface area for diffusive maternal-fetal exchange. The total surface of the villi at term is usually around 11-16m^2 [48, 63]. Each type of chorionic villus consists of an outer epithelial trophoblast layer and a vascularised stromal cell core.

The *mesenchymal villus* is the precursor to all other villus types [56]. They develop from 5-23 weeks of gestation and are infrequently seen at term [61, 64].

The *immature intermediate villus* is the first villus type to differentiate from mesenchymal villi [56]. Present from 8 weeks of gestation, by week 20 they have differentiated into the structural stem villi and are essential for establishing the structural support of the villus trees. If seen in large quantities at term, they indicate a lack of villus maturation [65].

The *stem villus* is the base (or trunk) of the villus tree providing structural support and vascular transport [23, 56]. They are formed from the immature intermediate villi after 18 weeks and remain until birth. Large stem villi support a single fetal artery and vein but may also contain smaller vessels at their periphery.

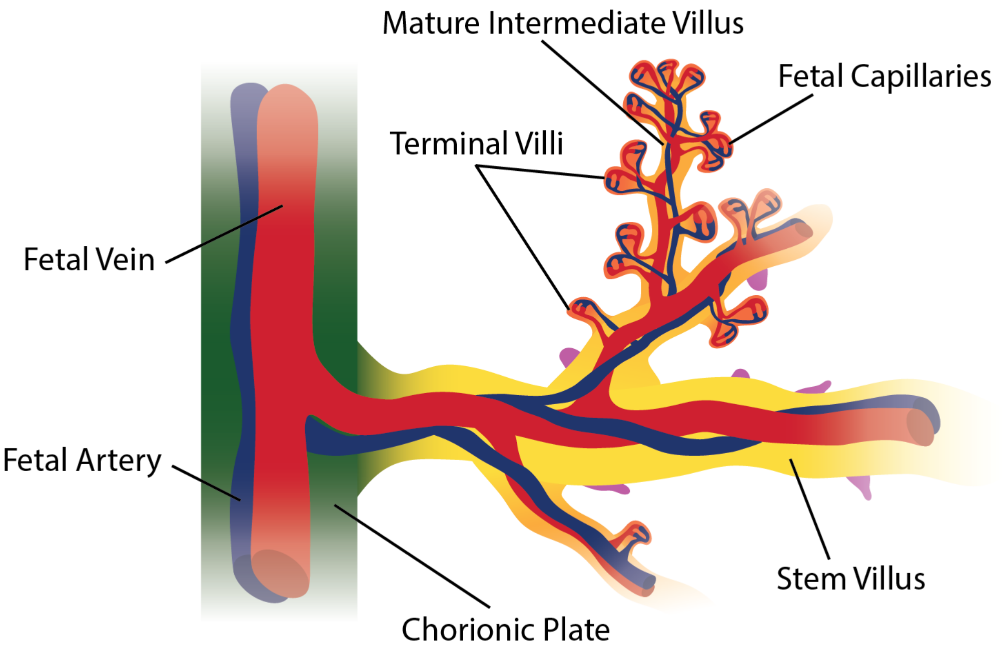


Figure 2.2: Fetal vascular system of the term chorionic villi. The fetal umbilical vein and arteries connect to the chorionic plate from which the chorionic villi grow. Stem villi provide a structural trunk and vessel transport for the rest of the villus tree. From the mature intermediate villi grow the capillary-rich terminal villi with low diffusion distances and high surface areas. Fetal blood is oxygenated by diffusive exchange at these capillaries before being transported back through the chorionic tree to the umbilical vein.

The *anchoring villus* is a stem villus which has come into contact with and anchored to the maternal stroma. By fusing against the basal plate, they maintain the structure of the rest of the villus tree and allow for the proliferation of trophoblasts into the maternal decidua, a process essential for further decidualization and remodelling of maternal tissue [66].

The *mature intermediate villus* is the second villus type to differentiate from mesenchymal villi, preferentially differentiating over immature intermediate villi from the 24th week [56]. Mature intermediate villi contain both small vessels and capillaries and are the primary site of diffusive exchange before terminal villi develop [64].

The *terminal villus* is the end stage of villus maturity, growing from the mature intermediate villi rather than differentiating from another villus type. They appear at 27 weeks and are the dominant villus type by the 34th week to term [56]. Terminal villi provide the shortest diffusion distance of 0.5-2 μm, more than 50% of their stromal volume is composed of capillaries, and there is a decelerated fetal blood flow of around 80% within these capillaries [61, 64]. A lack of terminal villi or too many small and sparsely separated terminal villi at term indicate a villus maturity disorder due to a hypoxic environment [64, 65].

The *villus sprout* is a protruding cluster of trophoblast cells growing from any other villus type. They are the first step in new villus branching growth in early gestation. At later gestation, they contain a clustering of syncytiotrophoblast nuclei (syncytial knots), which decrease the syncytium width, thereby optimising maternal-fetal diffusion [56].

Maternal/Fetal Surfaces

On either side of the placenta parenchyma are the maternal and fetal surfaces: the basal plate and the chorionic plate. These surfaces overlap and merge near the placental margin, called the marginal zone, to enclose the intervillous space [67].

The *chorionic plate* is the entry point for the fetal vasculature into the placenta. It is the supporting base of the chorionic villus tree system to which the umbilical cord attaches and contains the primary fetal veins and arteries. It is partially fused with the outer amnion membrane.

The *basal plate* is the entry point for the maternal vasculature into the placenta. It is the boundary between the fetal placenta and maternal tissue, providing access for the uterine glands, spiral arteries and uteroplacental veins into the intervillous space [67]. As the maternal decidual tissue becomes a part of the basal plate, it will contain a mix of cells of both maternal and fetal origin [68]. Basal septa are protrusions of the basal plate into the intervillous space, partially dividing it and the contained chorionic villi into lobes [67].

2.1.4 Cells

Cells are single- and multi-functional biological primitives comprising all structures in the body. There are an estimated 36 trillion cells in the adult human male and 28 trillion in the adult human female [69]. Different kinds of cells perform vastly different roles, reflected in their morphologies, endocrine and chemical activities, cell-to-cell interactions, and locations.

In the placenta, cells can be grouped into three main categories: the trophoblast cells, the mesenchymal-derived cells (also called villus stromal cells), and extravillus cells. Some of these cells, such as fibroblasts and vascular endothelial cells, are found throughout the body, and some are unique to the placenta. This categorisation is not exhaustive but follows the current scientific convention where possible and aims to summarise conceptual cellular function across the placenta.

Trophoblast Cells

Trophoblasts are a cell type uniquely found in the placenta. Believed to have originated through the incorporation of endogenous retrovirus genes [57–59], they are the first cells to differentiate from the fertilised egg [70]. During early pregnancy, trophoblasts form the initial barrier between the mother and fetus and are essential for implantation and interaction with the decidualized maternal uterus [71]. The exact subgroups of trophoblasts are unknown, and consensus has changed over time [71–73], but for our purposes, any trophoblast outside of a chorionic villus structure will be referred to as an extravillus trophoblast [56, 74].

The *syncytiotrophoblast* is a single multinucleated epithelial cell layer forming the syncytium. It is the surface membrane of all chorionic villus structures, creating a physical and biochemical barrier between fetal tissue and maternal blood and preventing the maternal immune system from detecting foreign fetal tissue and triggering an immune response [59]. The syncytiotrophoblast is the site of the placenta's highest metabolic and endocrine activity and secretes many of the hormones essential for maintaining pregnancy [56, 75]. They are unable to proliferate and are sustained by the fusion of cytotrophoblasts [58]. As pregnancy progresses, syncytiotrophoblast nuclei will group into clusters called syncytial knots.

The *cytotrophoblast*, found in a secondary layer below the syncytiotrophoblast [76], are precursor stem cells to the syncytiotrophoblast and extravillus trophoblasts [58, 72]. A full layer of cytotrophoblasts surrounds each villus during early pregnancy, but as pregnancy progresses, they become sparse [76]. As such, the proportion of cytotrophoblasts across a slide relative to gestational age can indicate accelerated or decelerated villus maturation [77]. Hypoxia induces cytotrophoblast proliferation [76], and excess cytotrophoblasts have been associated with pregnancy complications such as maternal diabetes and preeclampsia [78].

The *syncytial knot* is a clustering of four or more syncytiotrophoblast nuclei [79]. As pregnancy reaches term, syncytiotrophoblast nuclei form into syncytial knots, reducing the diffusion distance across the vasculosyncytial membranes that separate maternal blood and fetal capillaries [72]. They are indicators of villus maturity [79] with a lack of syncytial knots suggesting delayed maturation and a diagnosis of fetal vascular malperfusion [80]. Likewise, too many syncytial knots can suggest maternal vascular malperfusion and an adaptive response by the placenta [72, 81].

The *extravillus trophoblast* is any trophoblasts external to a chorionic villus [56]. From implantation to term, trophoblasts leave fetal villi and invade maternal uterine tissues [66, 71]. At various sites (uterine glands, veins, arteries), they remodel the tissue to facilitate nutrient exchange [66]. The most well-documented of these processes is the remodelling of maternal spiral arteries through a breakdown of the endothelium into low-resistance, high-capacity vessels facilitating blood flow into the placenta [71]. Abnormal invasion of trophoblasts is associated with pregnancy complications such as preeclampsia, fetal growth restriction, and excessive implantation [82]. Extravillus trophoblasts are also present in fibrin, cysts [83] and lesions such as placental infarcts [84], but are thought to be of no clinical significance in these regions. Extravillus trophoblasts have been found in maternal blood and organs prenatally and postpartum [85].

Mesenchymal-Derived Cells

The mesenchymal-derived cells, also known as villus stromal cells, are cells which are only found within the chorionic villus stroma and differentiate from mesenchymal cells. In early pregnancy, the villus stroma predominantly consists of undifferentiated mesenchymal cells [61]. As the pregnancy progresses, these cells take on specialised roles.

The *undifferentiated mesenchymal cell*, also called V-cells [86] or not otherwise specified (NOS) mesenchymal cell, are stem cells to all other cells found in the villus stroma. Found in the loose reticular stroma of early gestation chorionic villi, they are not frequently seen at term [56, 61].

The *fibroblast* is a structural cell type that produces the extracellular matrix and types I and types III collagen that maintains the villus stroma [56]. They are present throughout the body, but in the placenta, they can take on unique fibrocyte-like roles [87]. Fibroblasts differentiate into myofibroblasts in response to injury and myocytes to support blood vessels.

The *vascular endothelial cell* lines the inner surface of blood vessels and capillaries and is essential for new vessel development (angiogenesis). In response to vascular endothelial growth factor, vascular endothelial cells proliferate and migrate to the source of angiogenic signals, supporting new vessel formation at this site [88]. As a primary function of the placenta is vascular diffusive exchange, inappropriate capillary density relative to gestational age indicates a perfusive deficiency.

The *vascular myocyte*, or smooth muscle cell, is a structural cell type found in the media of blood vessels throughout the body. It supports the elastic property of these vessels through its contractile ability. In the placenta, fibroblasts differentiate into myofibroblasts that become vascular myocytes [56].

The *Hofbauer cell* is a macrophage uniquely found in the placenta of fetal origin [89]. Unlike macrophages found in other parts of the body, they are not believed to be part of an immune response. Instead, they are thought to coordinate endocrine activity, signalling other villus cell types and promoting the maturation and development of the placenta mesenchyme [55, 56, 89]. However, there is still uncertainty regarding the exact function and role of Hofbauer cells [55, 89]. Hofbauer cells undergo morphological changes across gestation, becoming difficult to distinguish from other stromal cells without immunohistochemical staining [89, 90].

Extravillus Cells

I use the name extravillus cells to refer to cells found outside of and not derived from the chorionic villi.

The *maternal decidual cell* is a decidualized endometrial stromal cell found in the basal plate of the placenta [61, 91]. Mesenchymal endometrial cells differentiate into decidual cells in the presence of elevated progesterone levels approximately 6 days after ovulation [91]. They are essential for controlling trophoblast invasion and further decidualization [92]. Defective decidualization is linked to infertility, recurrent miscarriage and placenta accreta [91, 93, 94].

The *leukocytes*, otherwise known as white blood cells, are immune cells that can be found throughout the body, including the placental tissues and fetal and maternal blood. In large quantities, they indicate an immune response and, in the placenta, are used to diagnose a host of chronic or acute inflammatory lesions, such as chorioamnionitis, villitis and premature rupture of fetal membranes [95, 96]. Leukocytes may be subcategorised into neutrophils, lymphocytes, and monocytes, including plasma cells, macrophages and natural killer cells.

Fetal red blood cells are oxygen-carrying cells in the blood of fetal vessels. At earlier gestations, they are nucleated, but this nucleation decreases across gestation such that, when present in large quantities at term, they indicate a pathology [97, 98].

The *amniotic epithelial cell* or amniocyte is a pluripotent stem cell derived from the amnion which surrounds the fetus and chorionic plate. In the placenta, it is found in the amnion lining of the umbilical cord and chorionic plate.

2.2 Clinical Placenta Histopathology

Histopathology is the study of disease at the microscopic level through the analysis of histology slides. A histology slide is a sample of an organ prepared, sliced, stained, and placed on a glass slide (Figure 2.1) for examination under a microscope by a pathologist. Slides may be digitised using microscope scanners and analysed on computers using specialised software. Pathologists are clinicians who specialise in histology analysis, and perinatal pathologists are pathologists who further specialise in the placenta. Histopathology allows us to inspect an organ at the cell level and is vital to many kinds of clinical diagnosis. For example, taking a biopsy of an organ to identify if a tumour is cancerous is a part of histopathology.

Histology slides are prepared by histology labs, and placenta slide preparation is similar to that of other organs. The sampled placenta tissue is first formalin-fixed to preserve the structure and prevent decay. The tissue is dehydrated with alcohol and cleared using xylene to remove liquids and colouration. Once the tissue is prepared, it is embedded in paraffin wax and sliced into 5 µm thickness slices ready to be placed on a glass histology slide. The slice is stained, usually with hematoxylin and eosin (H&E), and mounted on the glass slide. The slide is now ready to be analysed under a microscope or digitised.

For placenta histology, there are four kinds of slides: umbilical cord, maternal surface, maternal/fetal extraplacental membranes in a roll, and the parenchyma (main section). Guidelines recommend that a placenta submitted to microscopy should be sampled for histology at least once for the umbilical cord, once for the membrane roll, once for the maternal surface, and at two parenchyma sites, in addition to any lesions which are identifiable from the organ at the macroscopic level [19, 41, 61, 77]. Placenta samples are usually larger than other organs as they are collected non-invasively and take up the full width of a glass slide.

Placenta histopathology is an important part of the clinical investigation workflow. However, it is financially infeasible to submit placentas from all pregnancies to examination [4, 22], although this is sometimes advocated [41]. Between 20-40% of placentas from US and UK pregnancies qualify for submission to microscopic examination [3, 34–36]. There are guidelines for qualifying criteria [22, 31, 99, 100], but hospitals are free to follow their own submission systems [4]. The clinical pathway for submission starts from any adverse indications during the current or previous pregnancies. For example, suspected placenta accreta from a 1st trimester ultrasound will qualify the sample for submission. At delivery, the placenta is examined at the gross macroscopic level, and any irregularities or indications of injury will qualify the sample for microscopic examination. At microscopic examination, a perinatal pathologist, but sometimes a general surgical pathologist if one isn't available [29], will analyse the histology slides and create a pathology report [8]. This report details histological abnormalities, such as a villus maturity disorder, inflammatory response, or lesions. The information collected throughout this clinical process will inform clinical diagnosis and suggested care.

Historically, placenta pathology reporting has suffered from a lack of uniform diagnostic language [4, 19, 43] and miscommunication between pathologists and clinicians [30, 101, 102]. Over the last decade, efforts have been made to standardise placental pathology reporting and lesion classification, with the Amsterdam consensus [77] being the most widely adopted of these. This consensus, and many that followed [3, 22, 103], systematically group placental lesions and disease phenotypes into five main placental pathologies or patterns of injury: acute inflammation, chronic inflammation, fetal vascular malperfusion, maternal vascular malperfusion, and other. These pathologies are associated with a host of immediate and life-long adverse maternal and child health outcomes. Placenta pathology can inform immediate clinical treatment, intervention and management for long-term health outcomes, predicts recurrence risk in subsequent pregnancies, and explains causes of pregnancy loss [1–9].

2.3 Deep Learning

2.3.1 Overview

Deep learning is a subset of machine learning. At a high level, machine learning is the process of going from input data, an algorithm (or model) which fits to or learns that data for some desired output, and the application of that model to new data of interest. Whilst the inputs and outputs can be highly complex, these algorithms are simply learnt mathematical mappings from input to output. The 'learning' part here refers to the model's ability to automatically adjust and improve its parameters based on patterns and structures it identifies in the data.

Popular machine learning algorithms include logistic regression, Support Vector Machines, decision trees, and k-means clustering.

Deep learning models, also called neural networks, distinguish themselves from these approaches by their large parameter counts and training schemes, which, in theory, create universal function approximators that can learn the mapping for any task. While there are many variants, deep learning models fundamentally consist of a series of neurons (the width of the network) across multiple layers (the depth of the network), weights connecting those neurons with an additive bias, and a non-linear activation function applied to the output of each neuron.

Deep learning models learn the values of their weights and biases through iterative forward and backward passes across the layers in the network. The forward pass computes the output for a given input using the current state of the weights and calculates the loss by comparing against an expected output. The backward pass applies backpropagation [104], a method which uses the chain rule to compute the gradient of the loss and update the weights in a direction that minimises this loss. By minimising the loss, the network iteratively learns a better mapping from all inputs in the training data to expected outputs. If the mapping has learnt the general relationship between inputs and outputs, it will perform well on future data of interest. However, if the mapping is simply a memorisation of the inputs to their outputs, the network has overfit and will not perform well on future data. The goal of training deep neural networks is to learn the true relationships in the data, i.e. a picture of a cat has a cat in it regardless of the cat's background, direction or breed.

Originally inspired by neuronal electrical signalling in the nervous system, deep learning models began as feedforward single-layer perceptrons and were later extended to fully connected multi-layer perceptrons (MLP) [105]. The backpropagation algorithm [104] allowed these models to update their own weights, but it was the advancement in computing power, particularly with GPUs, that facilitated increased model complexity and significantly more powerful networks. Deep learning is a rapidly growing field that cannot be covered in its entirety here. This background will provide a high-level overview of the major advancements that have led to the current state-of-the-art across domains, with a focus on the models used in this thesis.

2.3.2 Computer Vision

Computer vision applies deep learning models to image data, extending the field of image processing, which traditionally relied on threshold- and texture-based techniques [106]. The field of computer vision was accelerated with the introduction of the Convolutional Neural Network (CNN) AlexNet [107] when it became state-of-the-art on the large benchmark dataset ImageNet [108] in 2012. CNNs and convolutional layers, first described theoretically in the 1980s [109, 110], showed early promise with LeNet in 1998 [111]. Convolutional layers use learnable filters in a sliding window over the image to create feature maps. These filters typically learn to detect specific features, such as edges or textures, allowing them to encode different image representations. The convolution operation slides each filter across the image and computes the dot product of the filter's values with the underlying pixels at each position. The resulting compressed feature maps can then be passed to further convolutional layers or MLPs to compute the output.

These days, computer vision is split into three main categories: image classification, object detection and segmentation. Image classification assigns an output to an entire image with notable models including VGG [112] and Resnet [113]. Object detection identifies objects' centroid coordinates or bounding boxes, often with classes, within an image. Notable models include RetinaNet [114] and YOLO [115]. Segmentation is split into semantic and instance segmentation, where all pixels in an image are classified into shared classes (semantic) or as instances of classes (instance) with notable models including U-Net [116] and Mask R-CNN [117]. All of these models use convolutional layers in their architectures.

2.3.3 Graph Neural Networks

In the last few years, Graph Neural Networks (GNNs) have taken the idea of convolutions across local image pixels and extended them to graph data. Graphs are defined in their simplest form as nodes and edges which connect those nodes, where nodes and edges can have associated features. We may consider the pixels in an image as a graph in a regular grid where each pixel is a node, the node feature is the pixel value, and nodes are connected to their neighbouring pixels. The analogue to convolution, therefore, is a neighbourhood aggregation (in this case, the dot product) of connected nodes. This idea of convolutions applied to graphs led to the Graph Convolutional Networks (GCNs) [118] and message-passing algorithms of today for graph data of any shape and structure.

The graph learning field is split into two categories: inductive and transductive. Transductive models use the topology of the entire graph and learn from labelled portions or features to infer over the unlabelled portion of the graph. They learn a separate embedding for each node in the graph and are, therefore, highly tailored to the training graph's specific structure and feature distribution. These models are performant but cannot be applied to new graphs or handle changes to the existing graph topology. Inductive models instead learn how to apply aggregation across a graph for a given task and thus can infer over new unseen graphs and graph topologies.

The most popular paradigm in graph deep learning is message-passing aggregation. The core idea is to learn new node representations which effectively capture the features and structure of the neighbourhood for each node in the graph. A final set of layers, usually an MLP, will then use these new node representations to predict the output of interest. There are many message-passing strategies, but they all broadly follow the same structure. First, a message is computed between all neighbouring nodes. All the neighbour messages for the given node are then aggregated, and the node's features are updated with this aggregated feature. Depending on the number of layers in the network, this process is repeated for the neighbours of the neighbours where each layer handles that respective neighbourhood. A 3-layer GNN, for example, will traverse 3 neighbour hops outwards, and the 3rd layer will handle message-passing for those 3-hop neighbours (Figure 2.3). Each step of message, aggregation, and update can be performed by simple operations like concatenation and mean or be learnt by a subsidiary neural network.

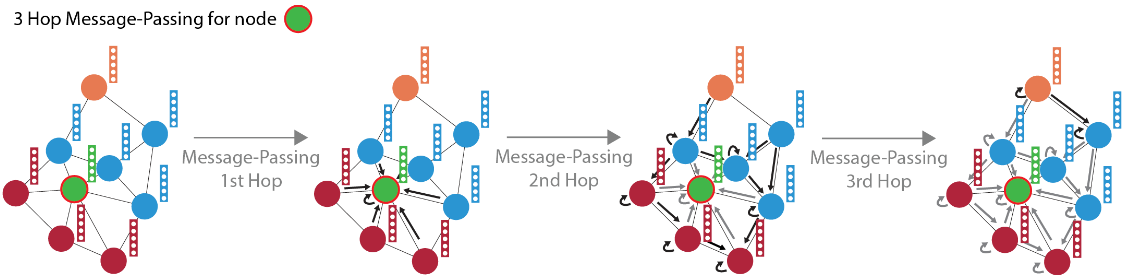


Figure 2.3: Message-passing for one node in a 3-layer graph neural network. For each node in the graph, the nodes up to 3 hops away will be iteratively aggregated using message-passing by a 3-layer graph neural network. Each layer in the network will aggregate that hop's neighbours to create a new node feature vector representing that neighbourhood.

There are three main learning tasks on graph data: node classification, link prediction, and graph classification. Node classification involves predicting the labels of nodes based on their features and structural neighbourhoods. It can also be used to cluster nodes into groups that share similar roles or neighbourhoods. Link prediction is the task of predicting edges between new nodes added to an existing graph, to unconnected nodes in a new graph, or to reconnect nodes for some new relationship. It is useful for recommendation systems [119] and protein interaction in biological networks [120, 121]. Graph classification involves compressing a graph into a single feature representation, which can then be used to predict a desired output for the whole graph, such as classifying chemical structure for drug discovery [122].

Graph neural networks tend to suffer from scalability issues. Most model architectures and training schemes require that the entire adjacency matrix and node features be available in device memory to the model during training. Graph sampling has emerged as a way to train across large graphs by loading only a portion of the graph into memory for a given batch. Sampling techniques include neighbourhood sampling [121], where all k-hop neighbours for a given node are loaded into memory, and subgraph sampling [123, 124], where the graph is pre-portioned into subgraphs and only those subgraphs are loaded into memory. The release of large benchmark datasets has facilitated the creation of newer scalable sampling techniques and software solutions [125–127].

2.3.4 Transformers

Whilst not a part of the work in this thesis, transformer models deserve mention as the current most popular architecture for some of the most powerful neural networks. Initially developed for natural language processing, transformers have been extended to many other domains, including images with vision transformers (ViT) [128] and graphs [129–131]. The key innovation of the transformer architecture is self-attention [132], a learnt importance of each part of an input to every other part of the input for the given output. In natural language processing, this would be the contextual relevance of every previous word in the sentence to the current word for the next word to be predicted. Unlike previous models in this domain, transformers do not process data sequentially and are, therefore, highly parallelizable. To account for the loss of sequential information, positional encodings are added to each input's features to reintroduce the data's original order. There is some evidence that convolution networks can perform as well as transformers with equal data and compute [133]. However, their ability to scale and distribute more easily across modern compute has made them the dominant architecture across a range of complex learning tasks on extensive datasets.

2.3.5 Types of Deep Learning

Across the different models and data modalities, deep learning algorithms can be split into two major types based on their training schemes: supervised or unsupervised. In supervised learning, models are trained using labelled data, where each input is paired with an expected output. The loss is computed based on how close the prediction is to the expectation, and the model is updated based on the magnitude of this error. Supervised learning was the dominant and highest-performing training scheme for some time but required curating a training dataset with representative, well-labelled examples and careful training to avoid overfitting.

In unsupervised learning, the model learns directly from the characteristics of the training data without any explicit labels. Autoencoders, for example, aim to reconstruct an input after it passes through an information bottleneck, usually via an encoder which compresses the data and a decoder that restores it to its original form [134]. They are useful for dimensionality reduction and the extraction of reduced features for downstream analysis. Generative Adversarial Networks (GANs) consist of two competing neural networks: a generator that produces fake examples of the input data and a discriminator that tries to differentiate between the two. Diffusion models, a popular alternative for image generation [135, 136], generate data by reversing a diffusion process to progressively denoise inputs [137].

Self-supervised learning is a subset of unsupervised learning where the model optimises towards a specific task using labels, just like in supervised learning, but those labels are derived from the input data. For example, in contrastive learning, the model learns to recognize similarities and differences within the data by optimising the distances between similar and dissimilar item pairs. These pairs are preselected algorithmically and are usually treated as binary labels. Self distillation models, like the student-teacher DINO [138] and DINOv2 [139], use outputs from earlier in the training process as labels for later training. Large language models (LLMs), like those using Generative Pretrained Transformers (GPTs), are trained to predict the known next word in a sentence after learning contextual relationships between words across large bodies of training text [140]. Self-supervised learning is competitive with, and in some fields surpassing, supervised learning. Without the requirement for manual labels, large datasets can be more easily prepared for training highly expressive models. Self-supervised models can be fine-tuned (trained with low learning rates) in a supervised setting for a specific task.

Foundation models have recently emerged in deep learning. Rather than being a new architecture, they are defined as very large models trained on vast quantities of unlabeled data, usually self-supervised, resulting in a model that can be adapted to a wide range of downstream tasks [141]. This model can be specific to a modality or domain but should apply to a wide range of diverse examples within that domain.

Examples of foundational models include LLMs like GPT-4 [142], image generation models like Stable Diffusion [135, 136], and task-specific models like the Segment Anything Model (SAM) [143]. Fine-tuning foundational models for specific use cases can be more powerful than training a model from scratch, and creating new foundational models across domains is an ongoing area of research.

2.4 Digital Pathology

Digital pathology, sometimes called computational pathology, is the automation of histopathology using image processing and machine learning techniques. It is used to replicate and assist in existing clinical workflows and to analyse histology data in the research setting [49]. Once a slide has been prepared for histology, it is digitised using a microscopy slide scanner to create a whole slide image (WSI) which can then be viewed with specialist software [144] or analysed programmatically.

Automating histopathology on WSIs is challenging for several reasons. Firstly, these images can be over 70GB per slide, making them infeasible to analyse as individual data points. To address this, a WSI is often split into smaller images called ‘tiles’ or ‘patches’, which can be analysed iteratively or in parallel. A second challenge is the variability in image features and visual appearance between slides prepared at different institutes and histology labs or even from the same lab across different days and operators [19, 145, 146]. This variability is caused by a lack of uniformity in staining procedures, such as stain concentrations and soak durations, as well as differences in slice thickness and slide scanners. Digital pathology methods must be invariant to these differences to generalise to future data. A final challenge is the noise inherent to this data domain. WSIs can have visual artefacts from slide preparation and scanning, 3D biological entities are arbitrarily sliced into a 2D view, and manual annotations or slide-level labels can contain a degree of mislabelling.

Historically, traditional image processing techniques were used for digital pathology [147]. However, these methods required manual parameter tuning, did not transfer well across datasets, and were not invariant to artefacts and noise. With the advent of deep learning, models can be trained with augmentations and cross-institute data to make them robust to slide variance [148, 149], artefact exclusion can be incorporated into model training and quality control [150, 151], and by training with large, diverse datasets, the true signal in the data can override noise and label incompleteness.

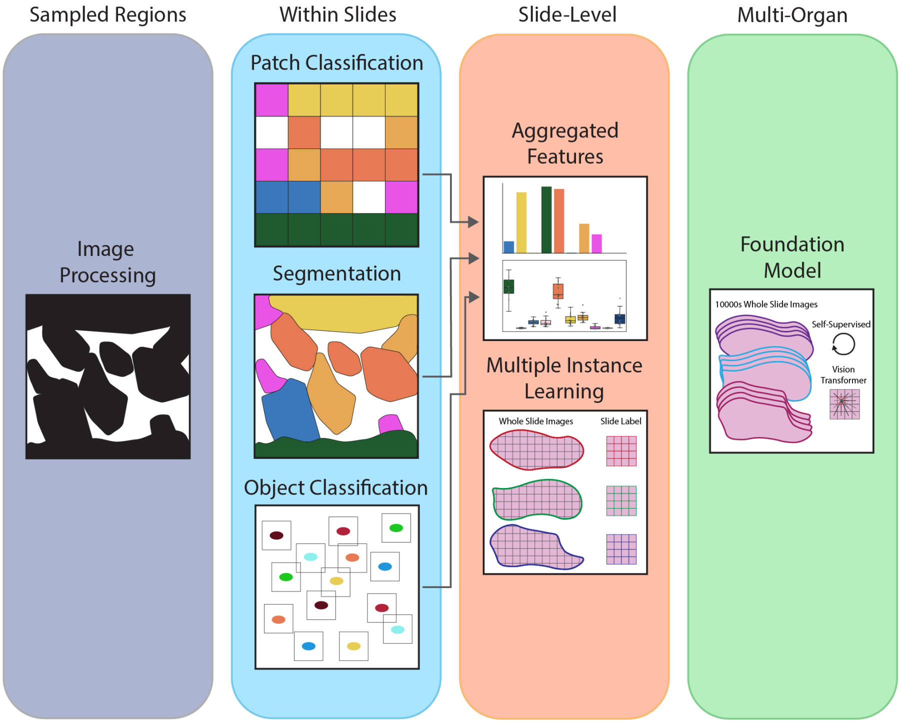


Figure 2.4: Approaches to digital pathology. Methods in digital pathology have ranged from image processing techniques within sampled regions of a slide to deep learning over patches or objects across a slide to predictions at the slide- or patient-level through feature aggregation or multiple instance learning. Foundation models have recently emerged as generalist models trained across large and diverse datasets for multiple tasks across many organs.

WSIs are ultimately a reflection of a single patient. The goal of digital pathology is to use a snapshot of biological data, like a histology slide, to make a prediction about the patient. Different approaches handle this in different ways. One approach is to identify relevant biological entities or features across a slide and aggregate that information into a slide- or patient-level prediction [152, 153]. Examples of these include nuclei coordinate localisation and cell classification [154], tissue segmentation [116], and patch image classification [152] (Figure 2.4). An alternative, which is particularly effective when only slide-level labels are available, is multiple instance learning (MIL). Here, a single datapoint, i.e. the slide, is treated as the composition of multiple instances of itself, often called a bag of instances, and this information is used as a prior when training a model. MIL is popular in digital pathology as slide-level labels are easier to extract from pathology reports, and instances can be patches or biological entities depending on the use case. Inductive GNNs, which are now being explored for digital pathology [155], are a kind of MIL where each node in the graph is an instance, and neighbourhoods encode local instance interaction [49]. Alternatively, vision transformers [128] for MIL treat patches like words in a sentence and use self-attention to capture long-range relationships between instances [49].

Digital pathology methods are usually developed for specific stain modalities, organs, and/or use cases. Recently, foundation models have emerged for cross-organ prediction across a range of tasks [156–161]. As with foundation models in other domains [141], these models are trained on large and diverse datasets using self-supervised learning [162] and vision transformers [128, 163, 164] and may be further adapted to specific use cases [165]. Whilst these models are not currently trained with data from all organ histology, they may provide good pretraining bases for transferring to the remaining organs and other stain modalities.

Ongoing digital pathology research is facilitated by open histology slide repositories (GTEX, TCGA, IDR, Big Picture), datasets (PanNuke [166, 167], Lizard [168], MedFM [169]), and grand challenges (CAMELYON17 [170], BACH [171], GlaS [172], OCELOT [173]). Much of the literature and model development use these open resources, supplementing with in-house data when available. This is perhaps one of the reasons (in addition to funding [45]) that most digital pathology is developed for cancer grading, subtyping, and survival prediction. Digital pathology has the potential to enhance clinical workflows by improving diagnostic accuracy and operational efficiency. It can accelerate research by offering high-throughput processing of slides and identifying new biomarkers for disease. It is beginning to enter the clinical space through commercial offerings with FDA or CE-IVDR approvals from companies such as Paige, Ibex and Visiopharm.

In everything there is a share of everything.

— Anaxagoras

3

Nuclei Localisation and Cell Classification

Contents

3.1 Introduction

Cells compose all structures in the body. In the placenta, higher-level features and functions can be inferred from the presence and proportion of specific cell types. For example, few cytotrophoblast cells and many syncytial knots indicate placental maturity, while the inverse indicates immaturity [61, 72, 76]. Likewise, the proportion of vascular endothelial cells found around capillaries and vessels can provide a proxy for the vascularity and capillary density of this highly vascular organ [72]. Deviations in these cellular proportions from expectations relative to gestational age can indicate pathological processes [72, 76, 82]. Similarly, an accumulation of leukocytes indicates an inflammatory response and can inform a diagnosis of chronic or acute inflammatory lesions [95, 96]. Understanding the cellular profile of a placenta slide allows us to summarise its function and serves as a good basis for hierarchical phenotyping.

The nucleus of a cell is one of the smallest and easiest structures to identify on a histology slide stained with the most common staining type, hematoxylin and eosin (H&E). The high DNA (and RNA) content in the nucleus has an affinity for the hematoxylin stain, which leads to the characteristic purple/blue colouration [174]. As nuclei are present in most cell types, we begin our bottom-up approach at the level of the nucleus, starting by identifying all nuclei across a placenta slide and then classifying them into cell types.

In this chapter, I present work towards building a fast whole slide-level nuclei localisation and cell classification supervised deep learning pipeline. The pipeline has three components: image processing, nucleus localisation, and cell classification. The image processing component first subsections and rescales the WSI into overlapping patches for the nucleus localisation component. The nucleus localisation component uses an object detection model to identify the nuclei within each patch. Using these nuclei coordinates, the image processing module crops patches around each nucleus and inputs them to the cell classification component. The cell classification component uses an image classification model to classify the nuclei as belonging to one of 11 placental cell types.

In the following sections, I motivate the 11 placental cell types included in this analysis. I summarise the literature in digital pathology for detecting nuclei and cell types and whether this has been applied to the placenta. Finally, I present and evaluate my deep learning methodology on parenchyma slides collected at three institutes for model performance and generalisability with a pilot exploration of model transfer to umbilical cord slices.[1]

3.2 Cellular Phenotypes of Interest

Following the categorisation presented in Chapter 2, I group cells in this analysis into three categories: the trophoblast cells, the mesenchymal-derived cells (also called villus stromal cells), and extravillus cells. Of the trophoblast cells, I include the syncytiotrophoblast, cytotrophoblast, syncytial knot, and extravillus trophoblast. The syncytiotrophoblast allows us to demarcate villus boundaries, cytotrophoblasts and syncytial knots are a measure of placental maturity, and the presence of extravillus trophoblasts outside of the maternal decidua can indicate lesions such as fibrin, cysts and placental infarcts [83, 84]. Of the mesenchymal-derived cells, I include undifferentiated mesenchymal cells, fibroblasts, vascular endothelial cells, vascular myocytes and Hofbauer cells. These are all of the cells normally found in the villus stroma, excluding pericytes, which are difficult to identify from H&E stains. Finally, of the extravillus cells, I include the maternal decidual cells and leukocytes. The maternal decidual cells can be used to identify normal structures not part of the chorionic villi, and leukocytes in large quantities indicate an inflammatory response. See Figure 3.1 for examples of these cells in H&E histology.

Cells not included in this analysis are those less commonly found in parenchyma slides of later gestation placenta histology and cells or cellular-level structures without nuclei. These include amniotic epithelial cells (or amniocytes) and endometrial epithelium/glands, which are of greater importance in membrane roll and maternal surface biopsy slides, and nucleated red blood cells, which are uncommon at later gestational ages.

---

[1] Sections titled *Related Work*, *Methodology*, *Test Data Performance*, and parts of *Domain Generalisation and Augmentations* have been adapted and expanded on from [175] (Nature Communications 2024). Sections titled *Cellular Phenotypes of Interest*, *A Pilot Case Study for Umbilical Cord Whole Slide Images*, and *Limitations* are new work.

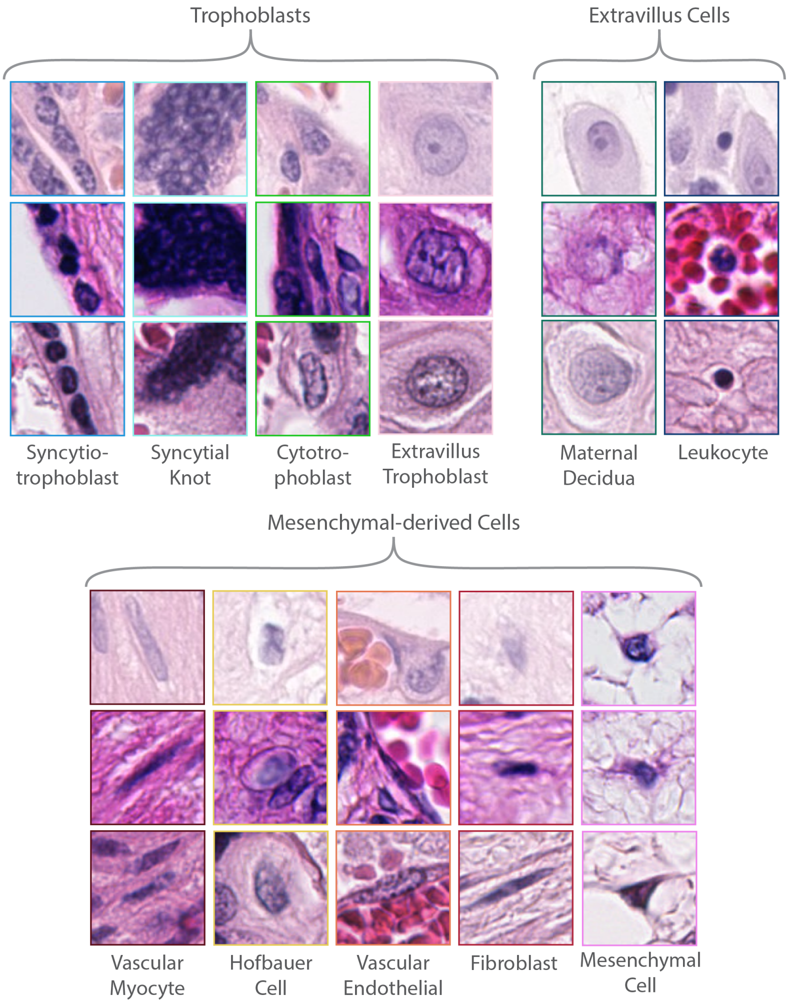


Figure 3.1: H&E-stained placenta cell types included in this analysis. Cells are grouped by category, and each row contains an example from a different institute (excluding mesenchymal cells, which are not present in slides from all institutes).

3.3 Related Work

Due to the high hematoxylin stain affinity of nuclei, traditional image processing techniques such as thresholding [176] and watershed algorithms [177] have been used to segment nuclei [144, 178, 179]. These methods are simple, effective and relatively computationally efficient but often require manual parameter tuning, especially on new data. In recent years, these methods have been superseded by deep learning techniques with the release of large and comprehensive open training datasets to support them. To name a few: the CoNSeP dataset has 24,319 nuclei in 4 categories from the colon [154], the MoNuSAC dataset has 46,909 nuclei in

4 categories from 4 organs [180], the PanNuke dataset has 205,343 nuclei in 5 categories from 19 organs [166, 167], and the Lizard dataset has 495,179 nuclei in 6 categories from the colon [168]. Popular deep learning models for nuclei localisation or segmentation and cell classification include HoVer-Net [154] and SONNET [181] among others [116, 165, 182].

None of the existing multi-organ nuclei localisation and cell classification models include placenta cells in their prediction. This is likely due to the lack of placenta examples in the large open-source multi-organ training datasets and the lack of any openly available placenta whole slide repositories. To the best of my knowledge, the only prior literature using deep learning for placenta cellular histology analysis comes from prior work in the group [50]. Here, authors use a RetinaNet [114] to localise nuclei in patches and an ensemble of image classification models to predict one of five cell types in chorionic villi.

The work in this chapter builds on the established cellular phenotyping methodology in the placenta through model improvements, training on new data across multiple institutes, expanding the number of predicted cell types, and by implementing an efficient inference pipeline for prediction across whole slides. The new nuclei localisation and cell classification models' performance in the placenta is comparable to the performance of popular models in other organs.

3.4 Methodology

3.4.1 Whole Slide Image Data

The placenta WSIs used throughout this research were obtained from clinical collaborators at three institutes. The placentas and their associated clinical data were collected as part of routine pathology investigation at each institute. From the first institute, the University of Tartu, Estonia, we have 110 placentas and 547 slides collected between 2016-2020 (hereafter referred to as UoT). From the second institute, the Hadassah Medical Center, Israel, we have 200 placentas and 831 slides collected between 2016-2017 (hereafter referred to as HMC). From the third institute, the Northshore University HealthSystem, Chicago, USA, we have 25 term placentas and 25 slides collected between 2021-2022 (hereafter referred to as NUH).

Histology slides were prepared using a standard formalin fixing, paraffin-embedded H&E staining procedure. As per clinical guidelines, appropriate, full-thickness sites were sampled, and 5 µm thickness slices were generated. Slides were scanned and digitised at x40 magnification using a Hamamatsu XR at UoT, a 3D HISTECH PANNORAMIC 250 Flash III at HMC and an Aperio GT 450 scanner at NUH.

3.4.2 Dataset Annotation

For training and evaluating the nuclei localisation and cell classification models, I manually annotated ground truth labels in nine randomly selected parenchyma slides from nine singleton pregnancies of 2nd trimester, preterm and term samples from both healthy placentas and those exhibiting a pathology from all three institutes. I annotated 16,883 nuclei in 252 image patches for the nucleus localisation model and 18,946 cells for the cell classification model (Figure 3.2). 64 image patches with no nuclei were included to discourage the nuclei model from predicting on the slide background. Annotations were randomly split into 70/15/15% train/validation/test datasets. The annotations were initially made in random regions across randomly selected slides, and to improve model robustness when training with a relatively small number of datapoints (<19k), datasets were bootstrapped by iteratively correcting prior models' predictions in regions of new, unseen slides (Figure 3.2). Annotations were made using QuPath v3.1 [144] and validated by a perinatal pathologist mentor in weekly meetings.

3.4.3 Whole Slide Image Patch Extraction

A single full-sized 25x75mm glass slide placenta parenchyma sample usually contains between 700,000 and 3 million cells. For efficient whole slide prediction of nuclei and cells, it is essential to have fast and parallelised reading and loading into memory of image patches extracted from the slide. Additionally, different brands of slide scanners output different proprietary slide file formats (e.g. .svs, .ndpi, .tiff) without

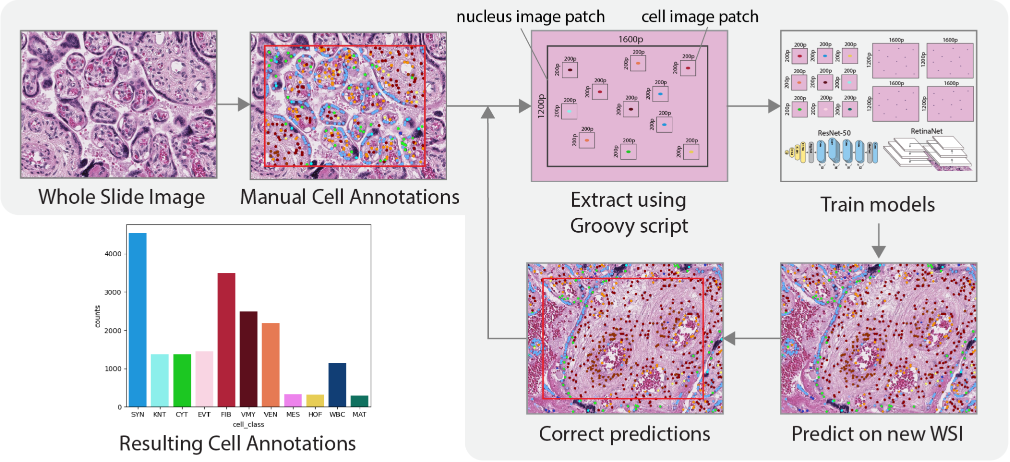


Figure 3.2: Nucleus and cell dataset annotation and bootstrapping workflow. Initial annotations are made using QuPath across slides, extracted using a Groovy script, and used to train the nuclei localisation and cell classification models. Predictions from these models are then assessed on new slides, and new annotations are made by correcting model predictions, which are then used to retrain the models. The resulting total manual annotated cell images, split by class, are shown in the bottom left.

a simple means of translation between them. To support most slide formats and for computationally efficient slide reading, I built an image processing module which selects between the slide reading libraries libvips [183] or BioFormats [184] depending on the file format. libvips is a C library for parallelising image processing with Python bindings in pyvips. It uses an underlying histology image reader C library, OpenSlide [185], with Python bindings in openslide-python. Bio-Formats is a histology image reading Java library that supports a large range of slide formats with Python bindings in python-bioformats. The module will preferentially choose libvips.

For the nuclei localisation stage, the module will partition the whole slide into 1600x1200 pixel (177.44x133.08 µm) patches, with an overlap of 200 (22.18 µm) to account for edge effects in model predictions at image boundaries. Patches with mean channel values >245 or <10 are removed to exclude patches containing no tissue, i.e. all white or all black images. Patches where the ratio of the mean 10% darkest portion of the image to the mean 10% lightest portion of the image is greater than 0.95 are also removed. This checks for a highly uniform, low-contrast image where the brightest and darkest parts of the image are similar in intensity. In practice, this removes images which lack significant visual content, usually because they are a uniform pale colour such as grey. For the cell classification stage, patches of 200x200 pixels (22.18x22.18 µm) are extracted centred on each previously predicted nucleus coordinate.

All patches are read and loaded onto devices (CPU or GPU) in memory without the need for additional on-disk storage to store extracted images. Given that slide scanners output WSIs with different pixel sizes per micrometre, to ensure that cells are consistently sized before being passed to the models, all patches are rescaled to 0.1109 micrometres per pixel. Metadata and results are stored efficiently in an SQLite database (Figure 3.3) with the ability to pause and continue partial inference across a WSI. 64-dimension cell embeddings from the cell classifier's penultimate layer are streamed to HDf5 files for downstream analysis.

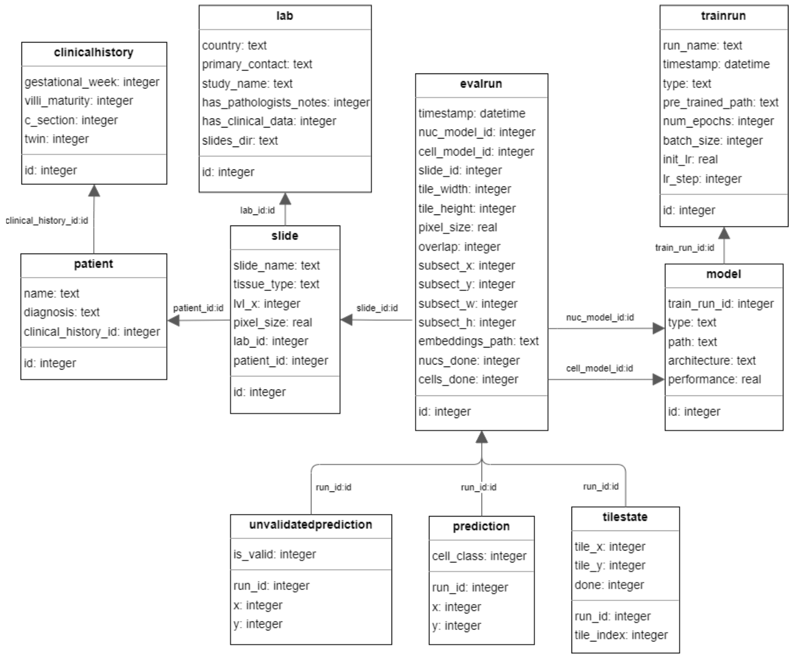


Figure 3.3: SQL schema for whole slide cell inference. This schema, implemented in an SQLite database, allows for efficient recording of cell predictions across slides with the ability to pause and continue incomplete runs. The schema centres around the table *evalrun*, which represents an inference pass across a whole slide using both the nuclei localisation and cell classification models, stored in *model.* Nucleus coordinate predictions are first saved in *unvalidatedprediction* before being processed into *prediction* with associated cell classes. The database contains relevant slide and patient information extracted from our clinical data, allowing for fast querying of slide characteristics. For example, the database can be queried for the number of slides (*slide*) for which the patient's placenta (*patient*) was described as having accelerated villus maturity (*clinicalhistory*) and whether any models (*model*) have run across those slides (*evalrun*).

3.4.4 Nuclei Localisation Training and Inference

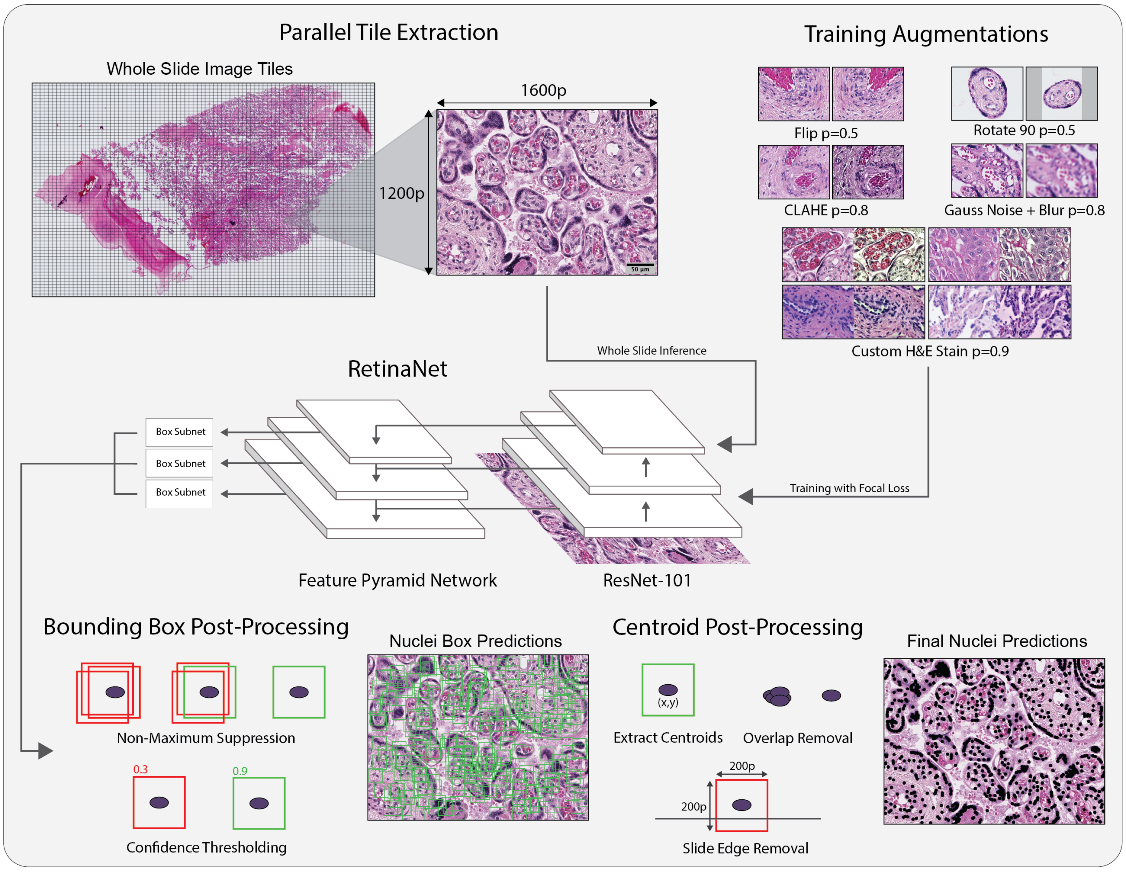


Figure 3.4: Nuclei localisation training and inference pipeline. The nuclei localisation model takes 1600x1200 (177.44x133.08 µm) pixel images extracted from the WSI with a 200 (22.18 µm) pixel overlap. Model bounding box predictions are first filtered with non-maximum suppression and a confidence threshold of 0.4 before extracting centroid coordinates. From the centroids, coordinates at the edge of the whole slide image and duplicate predictions caused by the overlap are removed. During training, images are subject to heavy image augmentation, including H&E-specific stain augmentations.

The nucleus localisation stage aims to predict a single centroid coordinate for each nucleus in a whole slide image. It takes as input 1600x1200 (177.44x133.08 µm) pixel images extracted from the WSI with a 200 (22.18 µm) pixel overlap, ensuring that all nuclei are shown whole to the model at least once. The model predicts initial bounding boxes, filtered using non-maximum suppression and a confidence threshold of 0.4 before centroid coordinates are extracted. From the centroids, coordinates at the edge of the whole slide image are removed, and any locations with duplicate nuclei generated due to the overlap are removed using a k-d tree [186] within a radius of 4 pixels. This radius threshold was chosen by finding, across a few slides, the nuclei to nuclei distance frequency peaks in a histogram of distances. At inference across a WSI using an NVIDIA A100 GPU and 6 CPU threads, the nucleus localisation stage detects ~1000 nuclei per second.

During training, a RetinaNet [114] with ResNet-101 [113] backbone is trained to predict bounding boxes around nuclei in the image, for which centroid coordinates are saved as the final prediction. The model is first fine-tuned from COCO [187] weights for 40 epochs, with an Adam [188] optimiser, ReLU activations, focal loss, and a 0.0001 learning rate with a 0.5 decay every 20 epochs. The model with the highest validation F1 score is then fully trained for 60 epochs with a 0.001 learning rate with the same hyperparameters, and the model with the highest validation F1 score is saved. Input images are subject to heavy image augmentation, including H&E-specific stain augmentations. On the validation and test datasets, model performance is evaluated using the F1 score of identified centroids within a certain distance (<3.3 µm) to the manually labelled ground truth points. This distance is smaller than typical nuclei radii and accounts for minor discrepancies from true centroids in the annotated data. See Figure 3.5 for a training and inference pipeline summary.

3.4.5 Cell Classification Training and Inference

After obtaining nuclei coordinates, the cell classification stage aims to predict the cell type to which each nucleus belongs. It takes as input 200x200 (22.18x22.18 µm) pixel images with each prior predicted nucleus at its centre and classifies these images into one of 11 placental cell types. As placental cell nuclei range in diameter from 5-20 µm, 22.18 µm^2 around each nucleus is large enough to capture the entirety of most cellular information in addition to contextual information surrounding the cell that may be relevant for prediction (i.e. red blood cells that are near vascular endothelial cells). As a final post-processing step, a k-d tree [186] is constructed across all syncytial knot predictions to convert isolated knots into the syncytiotrophoblast cell type and to group clusters of syncytial knot nuclei into a single point. Syncytial

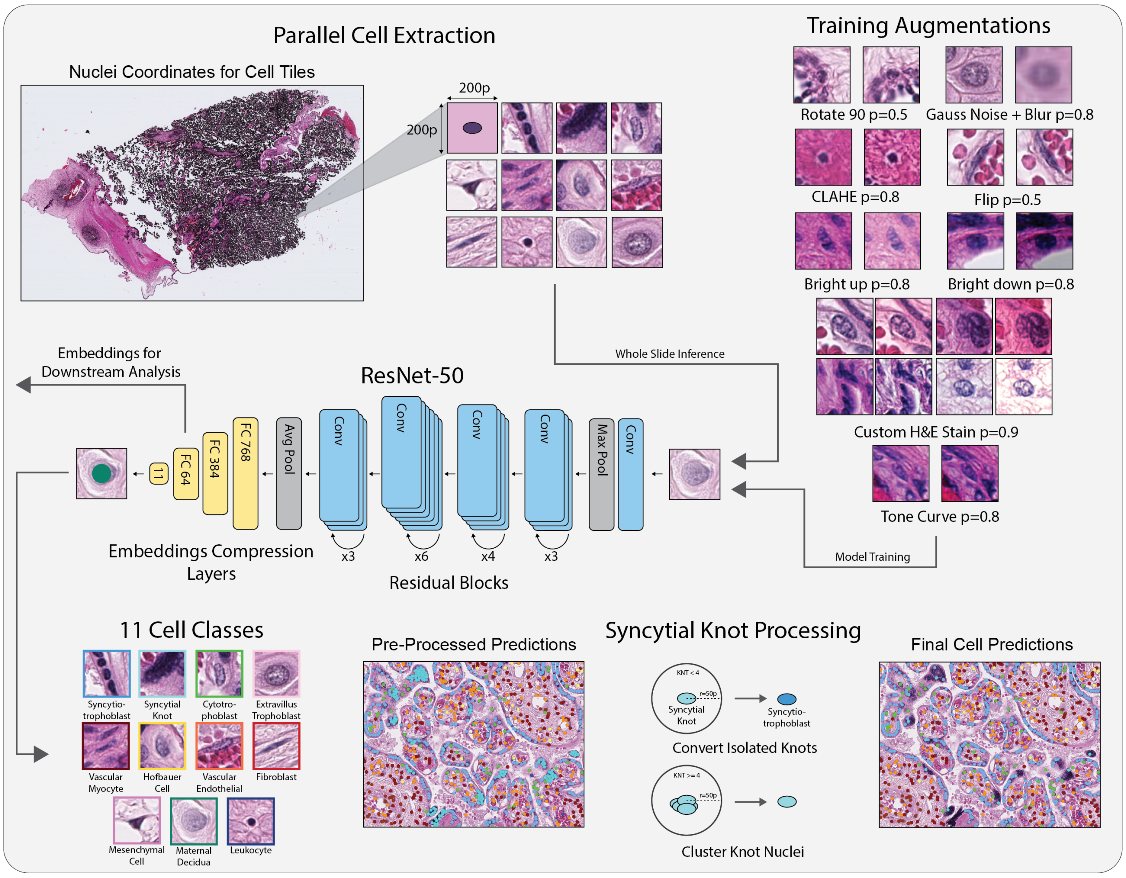


Figure 3.5: Cell classification training and inference pipeline. The cell classification model takes 200x200 (22.18x22.18 µm) pixel images extracted from the WSI centred around each nucleus coordinate and predicts one of 11 cell classes. Syncytial knots cell predictions with fewer than 4 cells within a radius of 50 pixels are converted into syncytiotrophoblast predictions, and those with 4 or more predictions are grouped into one. During training, images are subject to heavy image augmentation, including H&E-specific stain augmentations.

knot nuclei with fewer than 4 neighbours within a 50-pixel radius are relabelled as syncytiotrophoblast, and groups with 4 or more neighbours [79] have their neighbours removed. At inference across a WSI using an NVIDIA A100 GPU and 6 CPU threads, the cell classification stage classifies ~230 cells per second.

During training, a modified ResNet-50 [113] model with additional linear layers to reduce the latent vector dimensions to 64 is fine-tuned from ImageNet [108] weights for 60 epochs, with an Adam [188] optimiser, ReLU activations, cross entropy-loss, and a 0.0001 learning rate with a 0.5 decay every 20 epochs. The model with the highest validation accuracy is fully trained for 100 epochs with the same hyperparameters, and the model with the highest validation accuracy is saved.

Due to class imbalance, minority classes are oversampled during training to balance the distribution. As with the nuclei localisation model, input images are subject to heavy image augmentation, including various H&E-specific stain augmentations. See Figure 3.5 for a training and inference pipeline summary.

3.4.6 Augmentations

H&E stain intensity can vary between institutes for several reasons (stain concentrations and ratios, soak duration, slice thickness, different slide scanners) and can vary within an institute across days and operators, leading to batch-effects [19, 145, 146]. For models to robustly generalise to future data, they must be invariant to these stain differences. To improve model robustness, I train both the nuclei localisation and cell classification models with augmentations using the Ablumentations [189] library and custom H&E stain augmentations based on RGB to H&E colour conversion in [190]. Augmentations are described in Table 3.1.

Table 3.1: Training augmentations for nuclei localisation and cell classification. Augmentations are applied sequentially using the Albumentations library. StainAugment is a custom stain deconvolution method which varies hematoxylin and eosin (H&E) stain intensities randomly selected from one of 8 RGB to H&E colour matrices. Colour matrices are drawn from the original literature and derived from our own slides. *These augmentations were only applied to the cell classification model during training.

Augmentation	**Probability**	**Additional Parameters**
Flip	0.5	-
RandomRotate90	0.5	-
StainAugment	0.9	`variance=0.4`
CLAHE	0.7	`clip limit=3.0,` `tile_grid_size=(8,8)`
RandomToneCurve*	0.8	-

RandomBrightnessContrast*	0.8	`brightness_limit=(-0.1,0.2),`
		`contrast_limit=(0.0,0.0),`
		`brightness_by_max=False`

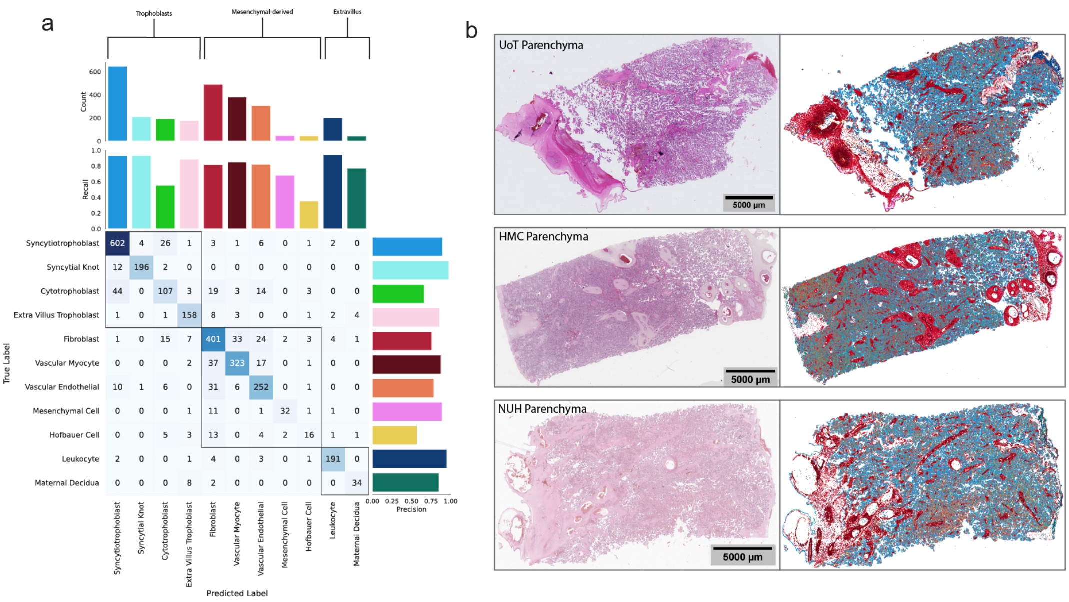
GaussNoise	0.8	`var_limit=(10.0,200.0)`
RandomToneCurve	0.8	`blur_limit=5`

3.5 Model Evaluation

3.5.1 Test Data Performance

On the test dataset, combined across three institutes, the nuclei localisation model achieves a 0.884 F1 score across 2,754 nuclei within 38 images. This is comparable to F1 scores reported by other state-of-the-art nuclei localisation models trained for other organs. HoVer-Net achieves an F1 score of 0.756 on the CoNSeP dataset [154] and 0.800 on the PanNuke dataset [166, 167] and SONNET [181] achieves an F1 score of 0.855 on the MoNuSAC dataset [180].

The cell classification model, evaluated across 2,743 cells for 11 placental cell types, achieves an overall accuracy of 84.29% and a top-2 accuracy of 94.90%, with a 0.9773 macro-averaged Receiver Operating Characteristic Area Under Curve (ROC AUC). Most model misclassifications are within closely related cell differentiation pathways (Figure 3.6). For example, the syncytiotrophoblast's next most predicted cell is the cytotrophoblast. Likewise, trophoblasts are confused with other trophoblasts and mesenchymal-derived cells with other mesenchymal-derived cells.



True Label \ Predicted Label	Syncytiotrophoblast	Syncytial Knot	Cytotrophoblast	Extra Villus Trophoblast	Fibroblast	Vascular Myocyte	Vascular Endothelial	Mesenchymal Cell	Hofbauer Cell	Leukocyte	Maternal Decidua
Syncytiotrophoblast	602	4	26	1	3	1	6	0	1	2	0
Syncytial Knot	12	196	2	0	0	0	0	0	0	0	0
Cytotrophoblast	44	0	107	3	19	3	14	0	3	0	0
Extra Villus Trophoblast	1	0	1	158	8	3	0	0	1	2	4
Fibroblast	1	0	15	7	401	33	24	2	3	4	1
Vascular Myocyte	0	0	0	2	37	323	17	0	1	0	0
Vascular Endothelial	10	1	6	0	31	6	252	0	1	0	0
Mesenchymal Cell	0	0	0	1	11	0	1	32	1	1	0
Hofbauer Cell	0	0	5	3	13	0	4	2	16	1	1
Leukocyte	2	0	0	1	4	0	3	0	1	191	0
Maternal Decidua	0	0	0	8	2	0	0	0	0	0	34

Figure 3.6: Cell classifier performance on unseen test data. (a) Confusion matrix and precision and recall values of cell classifier predictions. Cell types are clustered into categories (trophoblast, mesenchymal-derived, extravillus) and ordered by their test set counts. Categories are highlighted by the topmost brackets and squares within the confusion matrix. (b) Example healthy term WSIs and cell predictions from each institute.

3.5.2 Domain Generalisation and Augmentations

I explore how data augmentation, including our custom H&E stain augmentation, impacts the stain invariance of the nuclei localisation and cell classification models. For each institute, models are trained on data from one institute (seen data), with and without data augmentation, and evaluated on test data from the other two institutes (unseen data). The nuclei model is largely unaffected by differences across institutes with an average F1 score of 0.8746 on seen data and 0.8516 on unseen data (Δ -0.0230) when trained with augmentations (Table 3.2). In contrast, the nuclei model does not generalise as well when trained without augmentations within the same institute and on unseen institutes (F1 score 0.8593 vs 0.8242, Δ -0.0351). The cell model, when trained with augmentations, results in an ROC AUC value of 0.9693 on seen data and 0.8410 on unseen data (Δ -0.1283). However, when trained without augmentations, the 0.9722 AUC ROC value on seen data drops to 0.6903 on unseen data (Δ -0.2819), suggesting poor generalisability without data augmentation (Table 3.3).

I further investigate the impact of domain shift, where new data is substantially different from trained data [191], on the generalisability of the cell model. Upon visual inspection, the slides from NUH were substantially paler than those from the other two institutes. When I train a cell model on data from the other two institutes without the augmentation that increases and decreases brightness, the model only gets 30.4% accuracy on NUH test data. However, with the inclusion of this augmentation, the accuracy increases to 64.4%. This discrepancy shows a substantial domain shift, likely caused by the pale quality of the slides.

Assuming we do not have manual annotations to assess our model, we can also identify a domain shift by examining the biological plausibility of cell predictions across a slide. This mimics the practical application of deep learning digital pathology methods where one has access to trained models and new data but no manual labels. Should the cell predictions across a slide substantially shift from biological plausibilities, such as in Figure 3.7, we may assume there is a domain shift, and we need to retrain our models. This is a benefit of assessing a slide using a large number of independent cell classifications against a slide-level biological expectation, leading to a natively explainable AI approach. Once a domain shift has been identified, incorporating a small number of images (n=1,691) from the new institute into the training data allows the cell model to perform equally well across all three institutes with test accuracies of 0.8395, 0.8338, and 0.8757.

Table 3.2: The effect of stain augmentation on nuclei detection model generalisability. Nuclei detection model F1 score when trained with and without augmentations on training data from one institute (seen institute) and evaluated on test data from the two other institutes (unseen). All F1 scores are recorded on test data. Pairs of rows show corresponding results with and without augmentations, and the best result in each category and pair is highlighted in bold.

Model Type	Seen Institute	Unseen 1	Unseen 2
UoT with augs	**0.895 UoT**	0.868 HMC	**0.883 NUH**
UoT without augs	0.891 UoT	**0.883 HMC**	0.823 NUH
HMC with augs	**0.872 HMC**	0.851 UoT	**0.873 NUH**
HMC without augs	0.856 HMC	**0.860 UoT**	0.836 NUH
NUH with augs	**0.857 NUH**	**0.811 UoT**	**0.824 HMC**
NUH without augs	0.831 NUH	0.748 UoT	0.795 HMC

Table 3.3: The effect of stain augmentation on cell classification model generalisability. Cell classification ROC AUC when trained with and without augmentations on training data from one institute (seen institute) and evaluated on test data from the two other institutes (unseen). All ROC AUC scores are recorded on test data. Pairs of rows show corresponding results with and without augmentations, and the best result in each category and pair is highlighted in bold.

Model Type	Seen Institute	Unseen 1	Unseen 2
UoT with augs	0.978 UoT	**0.871 HMC**	**0.844 NUH**
UoT without augs	**0.981 UoT**	0.791 HMC	0.720 NUH
HMC with augs	0.971 HMC	**0.857 UoT**	**0.875 NUH**
HMC without augs	**0.972 HMC**	0.794 UoT	0.697 NUH
NUH with augs	0.959 NUH	**0.791 UoT**	**0.808 HMC**
NUH without augs	**0.964 NUH**	0.544 UoT	0.596 HMC

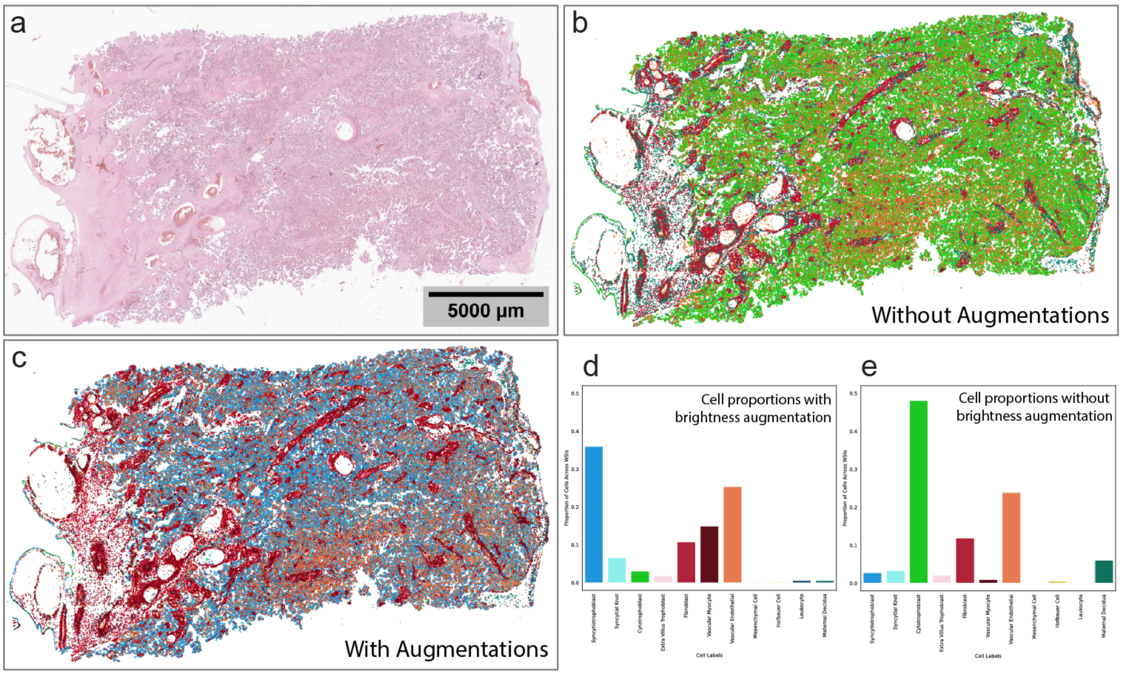


Figure 3.7: The effect of including brightness augmentations during training on cell predictions across a slide from an unseen institute. The pale quality of WSIs from this institute (a) leads to poor model generalisability when trained without brightness increasing and decreasing augmentations (b). Model predictions are outside of biological plausibility and indicate a domain shift (e). Including brightness augmentations improves generalisability (c) and biologically plausible predictions (d). In both cases, models are not trained with data from this institute.

3.6 A Pilot Case Study for Umbilical Cord Whole Slide Images

So far, the methodology has been built for cellular analysis of whole slide images of placental parenchyma. However, as previously mentioned, there are three other kinds of placenta histology slices: the umbilical cord, membrane rolls, and maternal surface biopsies. Certain lesions and pathologies can only be diagnosed from these slides. To explore how the placenta parenchyma models transfer to these other slices, I applied the nuclei localisation and cell classification methodology to one healthy umbilical cord slide from HMC (Figure 3.8).

To evaluate the models and for fine-tuning, I annotated 3,883 nuclei and 3,880 cells across the slide, introducing epithelial cells and macrophages as new cell types. The parenchyma-trained nuclei localisation model gets an F1 score of 0.799 on this data, suggesting that the model is robust enough to be applied to other slices from the same institute without further training. The parenchyma-trained cell classification model gets an accuracy of 61.5% on this data after mapping Hofbauer cell predictions to macrophage predictions, indicating a need for further training.

I trained a new model from the parenchyma cell classification weights with the new annotations split into 70%/15%/15% training/validation/test sets. The model was trained using the same training scheme as for the parenchyma. This new model gets an accuracy of 92.6% on test data, suggesting that the parenchyma-trained model weights can be a good pre-training basis for fine-tuning models on cell types specific to other slices.

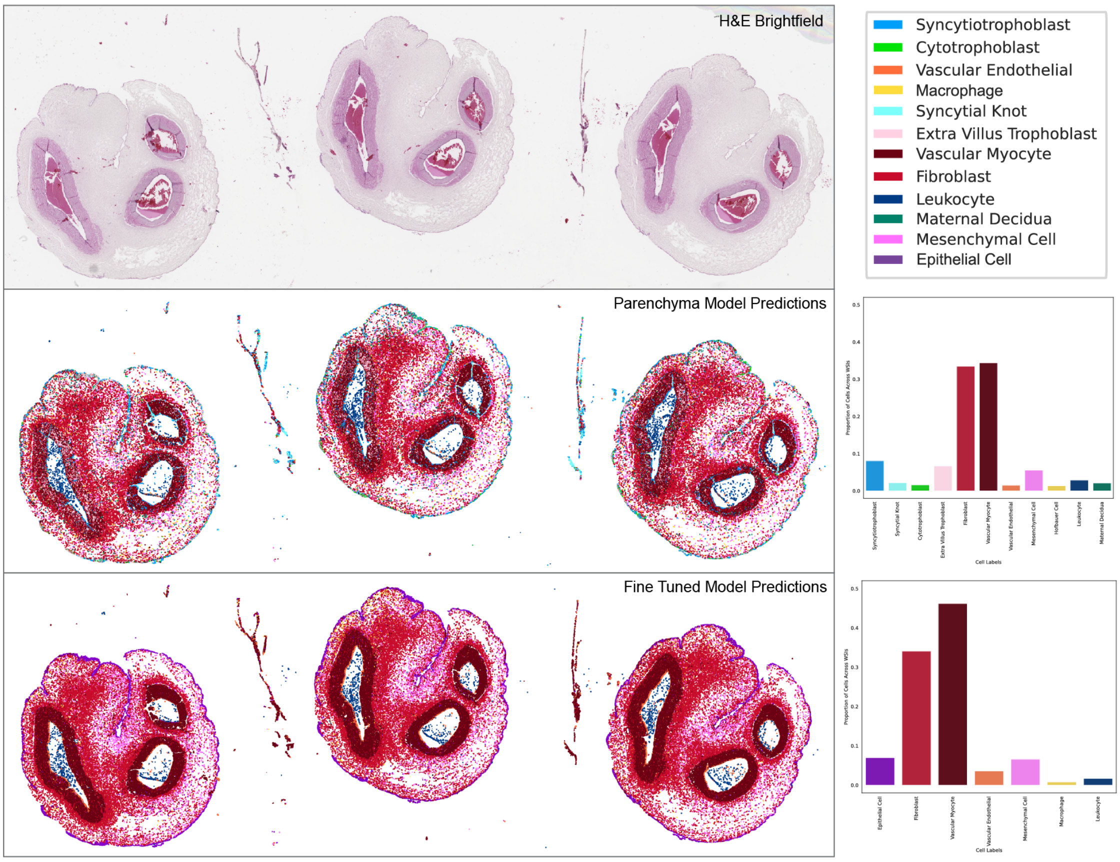


Figure 3.8: Healthy term umbilical cord cell predictions before and after fine-tuning. A healthy term umbilical cord whole slide image from HMC followed by the cell predictions of the parenchyma-trained model and the model fine-tuned with umbilical cord annotations. The proportion of each model's cell predictions across the slide is shown to the right of the image.

3.7 Conclusion

In this chapter, I have presented a deep learning pipeline for localising nuclei and classifying cell types within placental parenchyma whole slide images. This pipeline leverages an efficient image processing module for scaling to whole slide analysis, a nuclei localisation model with competitive F1 scores, and a cell classification model capable of distinguishing 11 distinct placental cell types with high accuracy. I show how training with augmentations, notably H&E stain augmentations, improves the generalisability of both models to data from unseen institutes and how a domain shift can be readily identified by comparing slide-level predictions against biological expectations. I explore model transferability to other placenta slide types by applying the models to an umbilical cord slide and show how the cell model can be fine-tuned to high performance using a small amount of newly annotated data.

In the context of the literature, this is the first work for automated placental cellular phenotyping across slides and the first method for classifying 11 different types of cells found in the placenta parenchyma. Many placental pathologies are associated with changes to the proportion or presence of these cells, and as higher-order structures are composed of them, they provide a good basis for further hierarchical analysis. This work facilitates a detailed analysis of the cellular composition and, by extension, the functional state of the placenta.

3.8 Limitations

As previously mentioned, the 11 cell types classified by the cell model do not encompass all possible cell types in the placenta. Whilst they provide a good interpretation of the function of any one placenta parenchyma, a more comprehensive analysis would include all cells found in the remaining three placenta histology slide types: the umbilical cord, membrane rolls, and maternal surface. Some additional substances found at the cell level may be useful to analyse. These include meconium, squamous debris, Wharton's jelly, fibrin/fibrinoid, bacteria and karyorrhexis. For some of these, like fibrin and Wharton's jelly, cellular community information may be sufficient to infer the presence of these regions. Higher quality cell labels may also be obtained using immunohistochemical staining for those cells on serially sectioned slides.

A downside of the proposed bottom-up approach is the reliance on nuclei as the smallest unit. Present methods will not detect any region containing biological tissue without nuclei. However, a simple preprocessing low-resolution thresholding technique, such as Otsu's method [176], which is commonly used for this purpose in digital pathology [192], would alleviate these concerns. The current methodology does not detect cells without a nucleus, such as red blood cells, but this may also be solved with thresholding. As is currently the case for syncytial knot nuclei, multinucleated cells will likely be predicted as multiple individual cells by the nucleus localisation model. However, as shown for syncytial knot nuclei, multiple predictions can be combined into one by using the predicted cell type for these nuclei and prior biological knowledge.

Finally, true cross-institute generalisability of models, without the need for manual annotation and retraining, remains a challenge. In practice, as with any application of AI models to new data, one will have to assess the performance of existing models and be aware of the possibility of a domain shift.

Yes, we have a soul,
but it's made of lots of tiny robots.

— Giulio Giorello

4

Hierarchical Tissue Microstructure Classification

Contents

4.1 Introduction

Identifying tissue microstructures is an important next step in modelling placental biology. The placenta is a rapidly changing organ, and the proportion and types of tissue microstructures across a slide indicate placental maturity and development [61, 81]. Deviations from the expected proportion of chorionic villus types relative to gestational age suggest accelerated or decelerated maturation and likely an adaptive response by the placenta [81]. Similarly, changes to the internal components of these microstructures are considered placental lesions and, in excess, are associated with poor pregnancy outcomes [77]. As an example, villi which lose their vasculature are termed avascular villi and indicate fetal vascular malperfusion [77, 80, 103] and villi with too much vasculature, such as capillary hyperplasia in chorangiosis, indicate a hypoxic environment [193, 194]. Tissue microstructures in the placenta are defined and distinguished from one another by their cellular composition [61], so a hierarchical cellular-based analysis provides a biologically explainable way of understanding these structures in both health and disease.

Assessing tissue microstructures in aggregate across a slide is an essential, if time-consuming, part of perinatal pathology assessment and reporting [3, 8, 195]. Clinical guidelines recommend that placental pathology reports state whether microstructures ‘correspond to gestational age’ and are ‘histologically normal’ [77]. A hierarchical cellular-based analysis, applied across whole slides at scale, can become an assistive tool for pathology reporting.

In this chapter, I present work towards building a hierarchical cell-based tissue classification framework based on the outputs of the nuclei localisation and cell classification pipeline. Taking the nuclei coordinates as nodes and the cell class embeddings as node features, I construct an Euclidean distance-based graph across a whole slide image. This whole slide cell graph is used to train an inductive graph neural network (GNN) for node classification of tissue microstructures. Effectively, after this step, each nucleus in a placenta whole slide image will have an associated cell and tissue microstructure class, and regions of tissue microstructures can be understood as the composition of their cells.

In the following sections, I motivate the 9 placental tissue microstructure types included in this analysis. I summarise the literature in digital pathology for hierarchical prediction of higher-order structures using cell graphs and GNNs, hierarchical patch-based approaches, non-hierarchical approaches, and any existing automated microstructure prediction in the placenta. I present and evaluate my deep learning methodology on parenchyma slides for model performance on test data and generalisability on data from a held-out institute. I perform an extensive hyperparameter exploration and replicate an existing unsupervised hierarchical benchmark from the literature. Finally, I show how all current scalable GNN architectures, including state-of-the-art, perform equally well on this data and describe a dataset release to the GNN community to motivate training on histology cell graphs.[1]

4.2 Tissue Microstructure Phenotypes of Interest

Following the categorisation presented in Chapter 2, I group tissue microstructures in this analysis into three categories: the chorionic villi, the maternal/fetal surfaces, and pathologic indicators. Of the chorionic villi, I include the stem villi, anchoring villi, mature intermediate villi, terminal villi, and villus sprouts. These are the villi most commonly found in the term placenta, for which surface area reference ranges exist. This excludes the immature intermediate villi and mesenchymal villi found at earlier gestations. The maternal/fetal surfaces are composed of the chorionic plate and basal plate and septums, which are both included. Finally, while not strictly tissue microstructures, I choose to include two lesions, fibrin and avascular villi, which in large quantities are pathologic indicators [77, 197–199] and can be identified from the composition of the prior predicted cell types. See Figure 4.1 for examples of these tissue microstructures in H&E histology and a schematic

---

[1] Sections titled *Related Work*, *Methodology* and *Test Data Performance* have been adapted and expanded on from [175] (Nature Communications 2024). Sections titled *Methodology* and *A New Node Classification Dataset for Graph Learning in Histology* have been adapted and expanded on from [196] (Neurips 2022). Sections titled *Tissue Microstructure Phenotypes of Interest, Generalisability and Standardisation, Hyperparameter Exploration and Ablation, An Unsupervised Cellular Community Clustering Benchmark*, and *Limitations* are new work.

representation of their relationships in the placenta parenchyma. See Table 4.1 for morphological descriptions and expected cellular compositions of the chorionic villi.

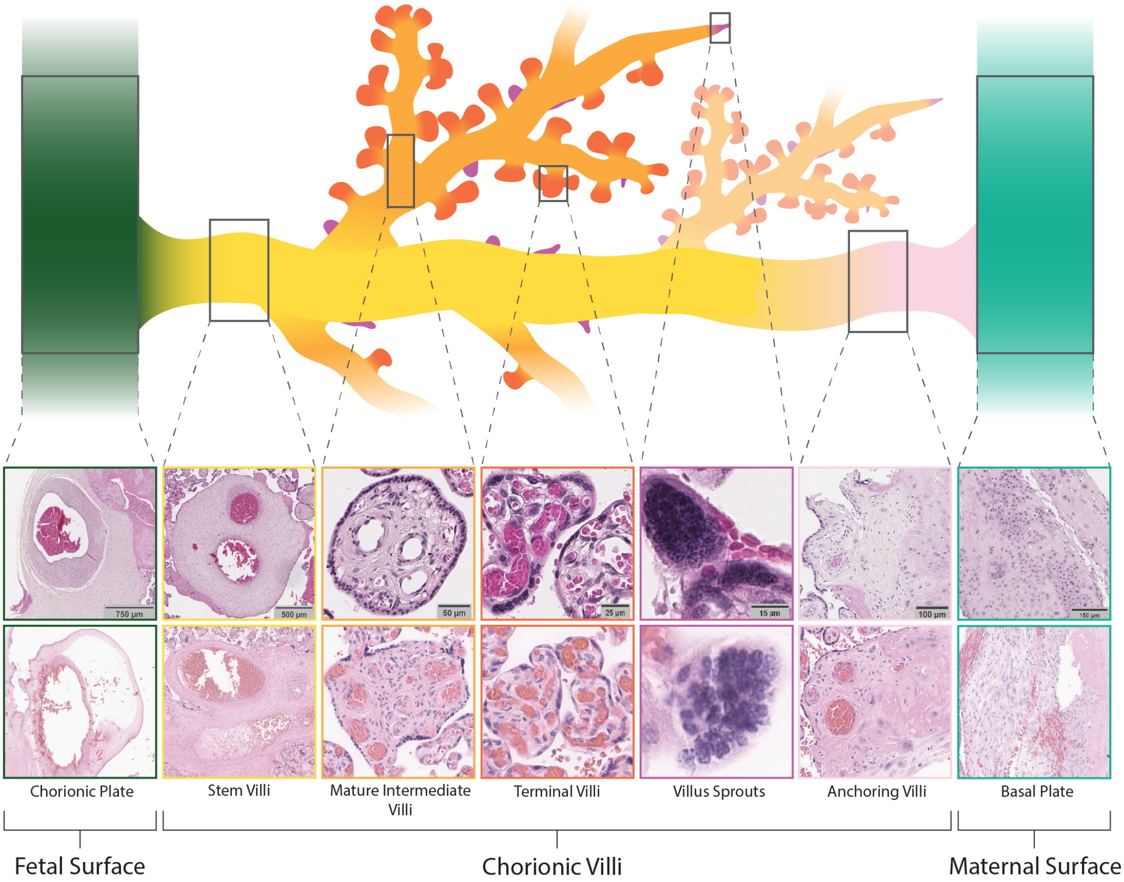


Figure 4.1: Schematic diagram with example images from H&E-stained histology of the healthy term tissue microstructures included in this analysis. Examples are drawn from slides at two different institutes. Not shown are avascular villi and fibrin, which are lesions and, in large enough quantities, suggest a pathologic process.

4.3 Related Work

Methods for hierarchically classifying higher-order structures in histology can be broadly split into patch-based or cell-based hierarchies. In patch-based hierarchies, the foundation is formed of image features extracted from a regular grid of patches across the whole slide image. In cell-based hierarchies, cells or the image features around cells form the foundation of the hierarchy and are usually represented as a cell graph.

Table 4.1: Characteristic and defining morphological and cellular traits of the chorionic villus tissue microstructures. The chorionic villi grow from one another in a branching, tree-like morphology. As such, the true underlying biology will reflect features across a continuum rather than discrete categories. However, the following traits are currently used to categorise the term chorionic villi. Table adapted from [61, 86]

Tissue Microstructure	**Diameter (µm)**	**Trophoblasts**	**Stroma**	**Vascularity**
Villus Sprout	60–100	Syncytiotrophoblast and cytotrophoblasts	None or reticular stroma	Unvascularised
Terminal Villus	60	Vasculosyncytial membranes, syncytial knots, syncytiotrophoblast	Scant fibrous stroma	Capillaries >50% of the stromal area
Mature Intermediate Villus	80–150	Complete syncytiotrophoblast, half cytotrophoblasts	Fibrous stroma	Capillaries <50% of the stromal area
Stem Villus	80–3000	Thick syncytiotrophoblast, frequently replaced by fibrinoid	Fibrous stroma	Large muscularized vessels, few capillaries
Anchoring Villus	80–3000	None or the same as stem villi	Fibrous stroma	Unvascularised or the same as stem villi

For nearly two decades, cell graphs have been proposed in biomedical literature as a means to identify and understand higher-order functional structures due to their inherently hierarchical composition of cells [200–204]. However, without automated means to apply these methods at scale, they used traditional image processing to identify cell coordinates, could not incorporate cell type in their analysis, and were limited to small regions within slides. More recently, leveraging the scalability of deep learning, cell graphs have been revisited for tissue phenotyping across whole slides where graphs and cellular features are extracted using deep learning models in histology [205–207] and spatial omics data [208].

With the advent of graph deep learning, in the last few years cell graphs and extracted cellular features have been used for training GNNs to discover new cellular phenotypes [209–211], make slide-level predictions [212–216], and hierarchically cluster cells into tissue microstructures or diseased regions [217–219]. However, for the most common imaging modality, H&E stained histology, these have only been applied to patch-level [214] and slide-level prediction [213–215, 220] or on graphs built within fixed subregions or patches [214, 217, 218]. Methods that restrict graph construction to cells within patches are sensitive to patch construction, lose information between patches, and cannot associate contextual information outside of a patch to information within the patch [213]. GNNs have also been applied to patch-based hierarchies, where graphs are built from a regular grid of patches across the whole slide and image features are extracted from each patch [221–226]. Overall, GNNs for histology analysis have succeeded in cancer grading, typing, and survival prediction [212–214, 216–226] and lesion classification [215, 227] with further reviews in [155, 228].

Patch-based hierarchical [164], or multiscale [229, 230] approaches which do not use GNNs have seen recent success with models trained on data from multiple resolutions of the same WSI. Additionally, although not designed specifically for tissue phenotyping, methods which use multiple instance learning for slide-level prediction can offer tissue structure localisation through the analysis of attention heatmaps on patches [192] or embedding similarity patch retrieval [158].

Non-hierarchical methods for classifying tissue structure in histology include deep learning image segmentation [172, 231–235] and patch classification [152, 235–238]. These methods have been less popular in recent years as segmentation models require manually annotated pixel-perfect masks to train, and patch classification limits prediction fidelity to the patch size [213]. These methods also have the same patch-based limitations mentioned previously, and as they are non-hierarchical, they do not capture or quantify the inherently hierarchical nature of biology.

In placenta histology, traditional non-hierarchical segmentation methods have been used to segment chorionic villi [26, 239–241] and capillaries [242]. A hierarchical deep learning method was developed for decidual vasculopathy localisation and severity classification [52]. Here, authors use an object detection model to localise all vessels across a slide, a binary classification model to label vessels as healthy or diseased and a random forest classifier on extracted vessel features and clinical delivery data to label the slide as healthy or diseased in aggregate. Automated analyses across other stain and imaging modalities for the placenta have been summarised in [243].

The work in this chapter describes a supervised hierarchical cell graph-based tissue microstructure prediction model for placental parenchyma histology. It is the first method for automated classification and localisation of tissue microstructures in placenta histology, one of few methods using GNNs on cell graphs built across entire whole slide images [213] and is the only method doing so for node classification. This approach provides a higher tissue prediction fidelity than patch classification, does not require precise manual annotation masks for segmentation, and does not restrict cell graphs to fixed patch sizes.

4.4 Methodology

4.4.1 Dataset Annotation

For training and evaluating the tissue classification model, I manually annotated ground truth labels in four randomly selected parenchyma slides from four healthy term singleton placentas from each institute. Annotations from two of these NUH slides were held out to evaluate model generalisability. I drew rough polygon boundaries around tissue microstructures on slides and used a Groovy script in QuPath v3.1 [144] to extract preloaded nuclei coordinates and assign them to the tissue classes of the polygons (Figure 4.2). This avoided the need to draw time-consuming segmentation masks and resulted in a higher prediction fidelity than a patch-based analysis. In total, I made 941,690 ground truth node labels, the polygons for which were validated by a perinatal pathologist mentor in weekly meetings.

Manual annotations are split into ~60%/20%/20% train/validation/test sets with 39,094 annotated nodes held out to assess generalisability. Rather than using random node splits, the validation and test regions from these labels are explicitly chosen such that they are larger than the 16-hop neighbourhood aggregation distance and contain a similar distribution of tissue microstructures to the training set (Figure 4.3). This contrasts with many graph learning node classification datasets,

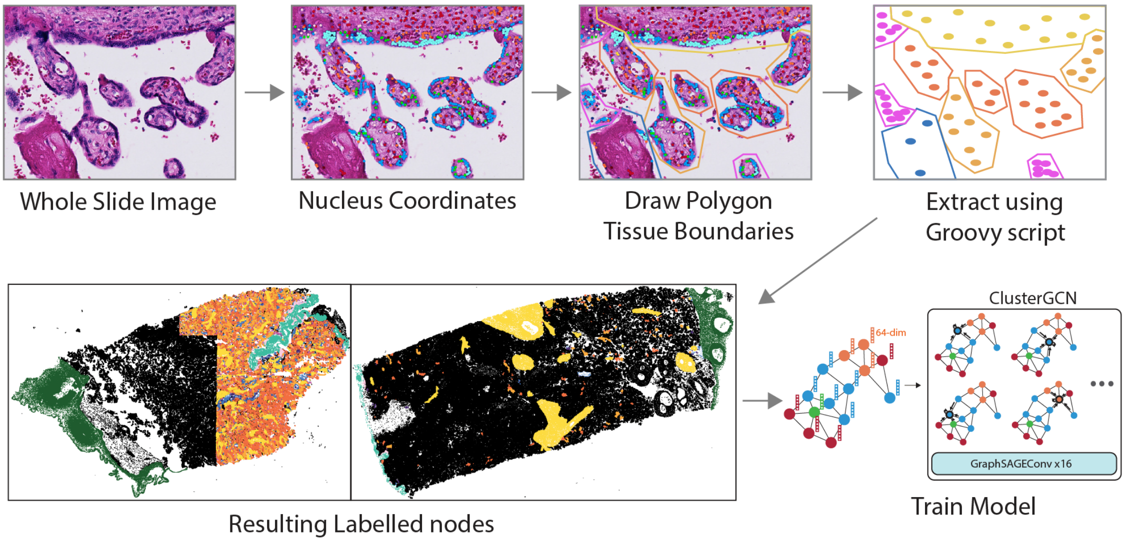


Figure 4.2: Tissue microstructure dataset annotation workflow. Previously predicted nuclei coordinates are first loaded across a whole slide image in QuPath. From these, the polygon tool is used to draw rough boundaries around same-type tissue microstructures. A Groovy script then extracts the coordinate of each nucleus and assigns it to the tissue microstructure type of the surrounding polygon. The bottom left panels show the resulting manually labelled nodes across two whole slide images used for training, with black nodes representing unlabelled nodes. These nodes are split into train/validation/test datasets and used to train the graph neural network.

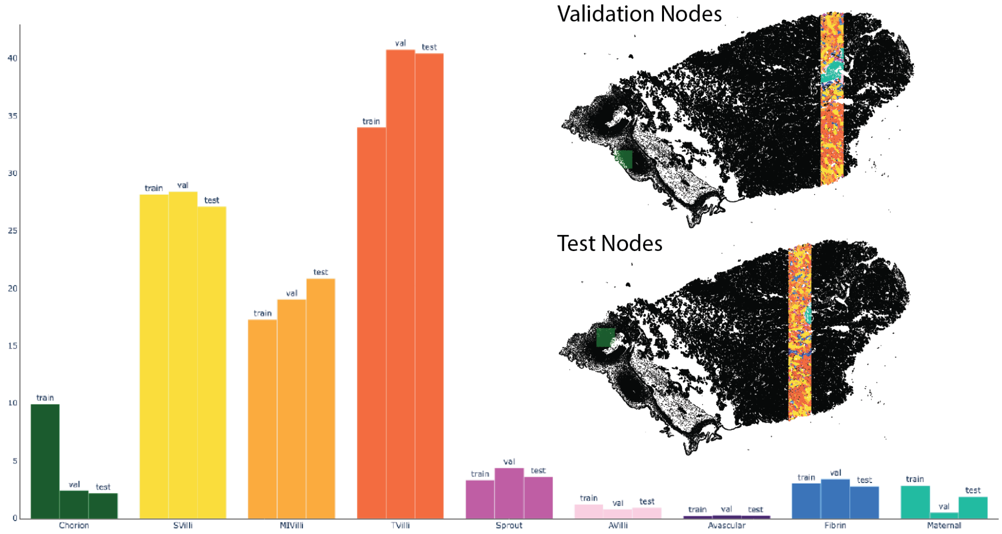


Figure 4.3: Tissue type proportions across dataset splits with validation and test regions on the slide. Validation and test regions are selected such that they have a similar distribution to the training set and are larger than the 16-hop neighbourhood aggregation distance.

which generate dataset splits randomly across nodes. In the context of cell graphs, nodes which share a neighbourhood will have similar aggregated vectors after message-passing, allowing a model to infer a validation node prediction from its neighbouring training node (Figure 4.4). As shown later, splitting datasets by region was found to avoid information leakage when different cells of the same tissues are shared across random dataset splits.

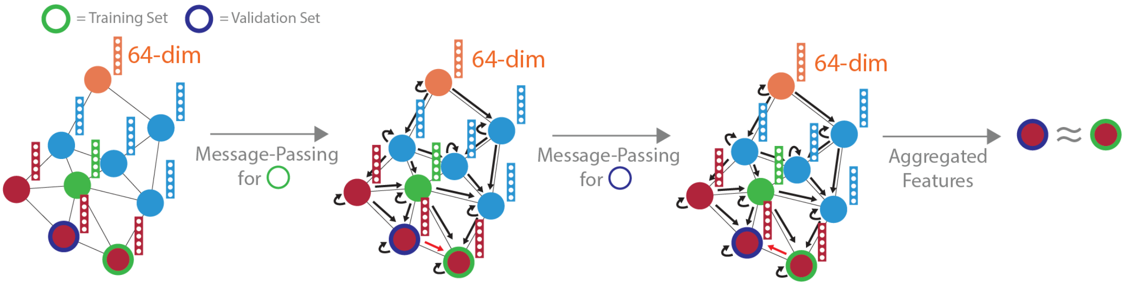


Figure 4.4: Aggregated features after message-passing for randomly split nodes. Randomly splitting nodes into datasets results in similar aggregated vectors after message-passing. This leads to information leakage across datasets and should not be used to train networks for node classification in cell graphs or any point cloud-style graph.

4.4.2 Cell Graph Construction

The tissue classification stage comprises cell graph construction and supervised graph neural network node classification. For the cell graph, I define nodes from the coordinate outputs of the nuclei localisation stage, with node features comprising 64-dimension embedding vectors from the penultimate layer of the cell classifier. The undirected, unweighted edges connecting the cell nodes are constructed from the intersection of two other edge-building algorithms, k-nearest neighbours (k=6) [244] and Delaunay Triangulation [245], with the addition of self-loops. At inference, k is increased to 8 for improved performance. The intersection edges benefit from the more sparsely connected Delaunay Triangulation graph but limit the number of edges which cross from one tissue boundary to another (Figure 4.5). This graph construction allows a message-passing model to aggregate cellular information within distinct tissue microstructures while accounting for differences in tissue size and internal cell distances.

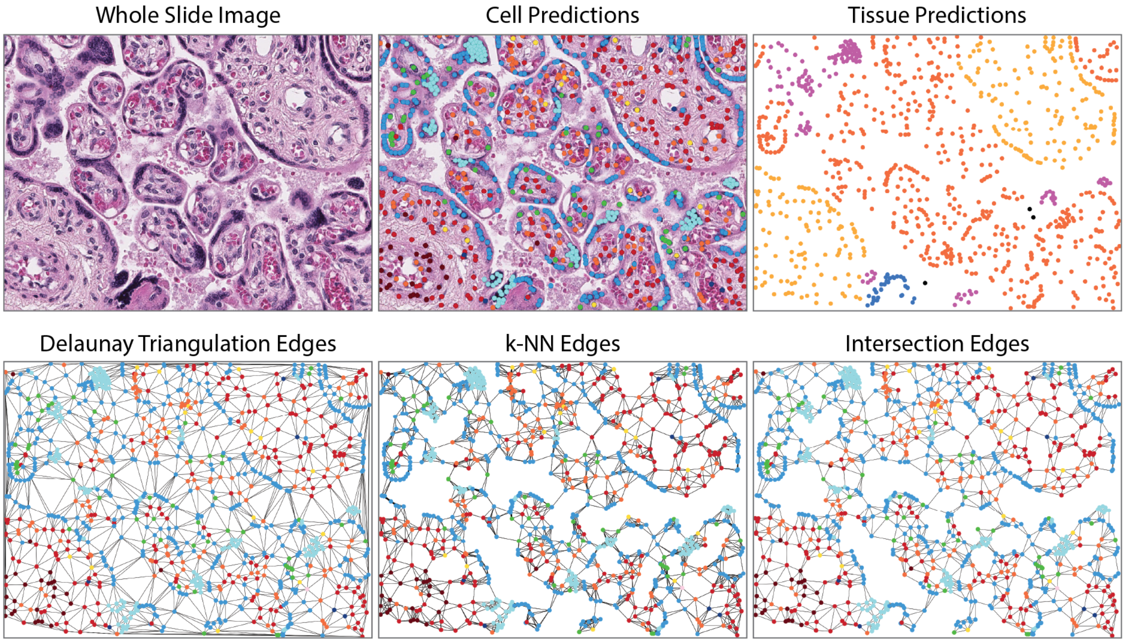


Figure 4.5: Resulting edges from three edge-building algorithms. Three different edge-building algorithms are explored using the whole slide nucleus coordinates and cell predictions. Delaunay Triangulation results in a fully connected graph with no overlapping edges but edges which jump across tissue boundaries. The k-NN edges provide more connections within tissue boundaries but result in redundant connections. The intersection edges, which take the intersection of both prior constructions, optimise for within-tissue boundary connections without overlap or redundancy.

For model training, I used the whole slide cell graphs from two whole slide images of healthy term placentas from UoT and HMC. Nodes within the graphs are assigned to training/validation/test set masks based on regions defined in an annotations file. The model predicts for each node in both graphs, and the loss is calculated only for nodes in the training mask. Whilst this is computationally less efficient, it is preferable to precropping dataset split subgraphs as the model can still use cellular node features of nodes outside of its dataset region, thereby avoiding edge effects.

Table 4.2 shows the graph characteristics for each of the two intersection edge construction cell graphs used for training. For *Homophily* and *Cell Homophily*, I calculate the node homophily [246], which measures the tendency for nodes to have the same class label or cell feature as their neighbours. It is calculated as the proportion of nodes whose majority of neighbours share the same value as the node itself, with 0 being no nodes and 1 being all nodes in the graph. We expect a high tissue microstructure label node homophily and a moderate to high cell node homophily. I additionally calculate the *Node Degree*, which measures the number of connections for each node averaged across the graph. Graph characteristics are calculated on the graphs as they would be input to the model, including undirected edges, which can double the edge count, and self-loops, which add the number of nodes to the edge count.

In Table 4.3, I contrast the two intersection cell graphs used for training against the k-NN and Delaunay Triangulation edge graphs. I show graph characteristics on the cell graph before making edges undirected and adding self-loops, after making edges undirected, and after adding self-loops to highlight the differences between the three methods. The intersection graph results in the smallest number of edges and node degree while increasing the label homophily and cell homophily, suggesting that edges are less likely to cross tissue boundaries.

Table 4.2: Graph characteristics of the two intersection cell graphs used for training GNNs. Characteristics are calculated after converting edges to undirected and adding self-loops. Homophily metrics are calculated using node homophily [246], with *Homophily* referring to tissue microstructure labels and *Cell Homophily* referring to the cell type of the cell embedding node features.

	Nodes	**Edges**	**Labels**	**Homophily**	**Cell Homophily**	**Degree**
Graph 1	1,112,555	5,648,705	675,521	0.9867	0.6561	5.0772
Graph 2	1,167,619	6,136,037	93,289	0.9971	0.6472	5.2552

4.4.3 Tissue Node Classification Training and Inference

I train a randomly initialised, inductive ClusterGCN [123] model with 16 GraphSAGEConv [121] layers and one 64-dimension linear layer for extracting embeddings for downstream tasks. The ClusterGCN subgraph sampling strategy was chosen for its computational efficiency and comparable performance to GraphSAGE [121], as demonstrated in a later section. The model is trained with 256 hidden units for 2000 epochs with an Adam [188] optimiser, ReLU activations, custom weighted cross entropy-loss, a 0.001 learning rate, a 0.2 node dropout, a batch size of 200

Table 4.3: Graph characteristics of cell graphs using all three edge-building algorithms. Characteristics are shown for original directed edges without self-loops, after converting to undirected, and after adding self-loops. Homophily metrics are calculated using node homophily [246] with *Homophily* referring to tissue microstructure labels and *Cell Homophily* referring to the cell type of the cell embedding node features.

Graph Type	Edges	Homophily	Cell Homophily	Degree
Directed edges without self-loops				
Intersection 1	2,268,075	0.9468	0.5606	2.0386
Intersection 2	2,484,209	0.9701	0.5573	2.1276
k-NN 2	6,675,330	0.9829	0.5478	6.0000
k-NN 2	7,005,714	0.9963	0.5493	6.0000
Delaunay 2	3,337,633	0.9536	0.5367	3.0000
Delaunay 2	3,502,828	0.9925	0.5342	3.0000
Undirected edges without self-loops				
Intersection 1	4,536,150	0.9823	0.5677	4.0772
Intersection 2	4,968,418	0.9961	0.5611	4.2552
k-NN 2	7,995,006	0.9783	0.5269	7.1862
k-NN 2	8,307,820	0.9958	0.5300	7.1152
Delaunay 2	6,675,266	0.9520	0.5291	5.9999
Delaunay 2	7,005,656	0.9921	0.5292	6.0000
Undirected edges with self-loops				
Intersection 1	5,648,705	0.9867	0.6561	5.0772
Intersection 2	6,136,037	0.9971	0.6472	5.2552
k-NN 2	9,107,561	0.9810	0.5857	8.1862
k-NN 2	9,475,439	0.9963	0.5891	8.1152
Delaunay 2	7,787,821	0.9587	0.5980	6.9999
Delaunay 2	8,173,275	0.9932	0.5976	7.0000

with batch normalisation, and a subgraph sampling size of 400 neighbours. Graph normalisation is applied to each cell graph to standardise node embeddings. The model with the highest validation accuracy, calculated without neighbourhood subgraph sampling for improved performance, is saved as the final model. For each node, the message-passing algorithm samples and aggregates the node features of nearby nodes up to 16 edge connections away, and the model uses these aggregated node features to predict the tissue type of that node. This aggregation procedure

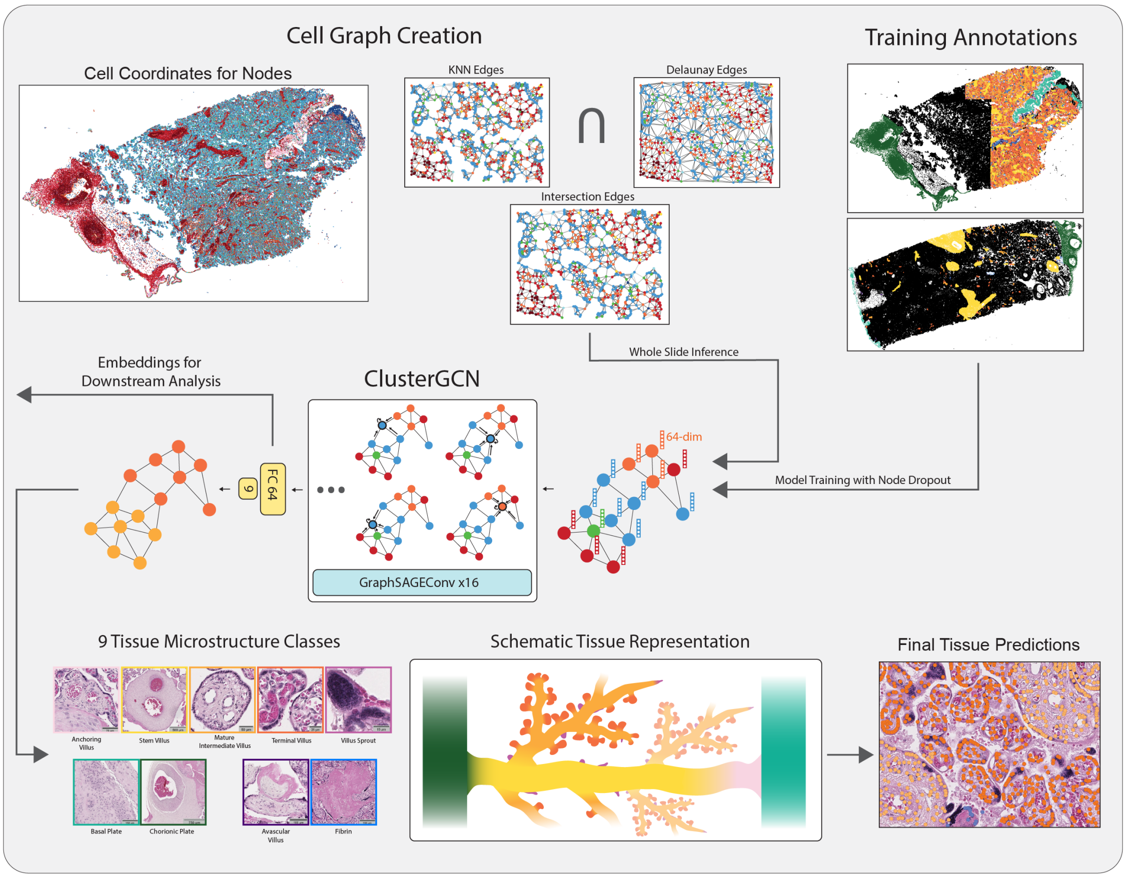


Figure 4.6: Tissue node classification training and inference pipeline. A cell graph is first built across the whole slide image using the nucleus coordinates as nodes, the cell 64-dim embedding vectors as node features, and the intersection of two edge-building algorithms as edges. A ClusterGCN model with 16 layers and additional linear layers for latent vector dimensionality reduction randomly subsamples subgraphs from the larger cell graph and predicts the tissue microstructure type of each node using message-passing.

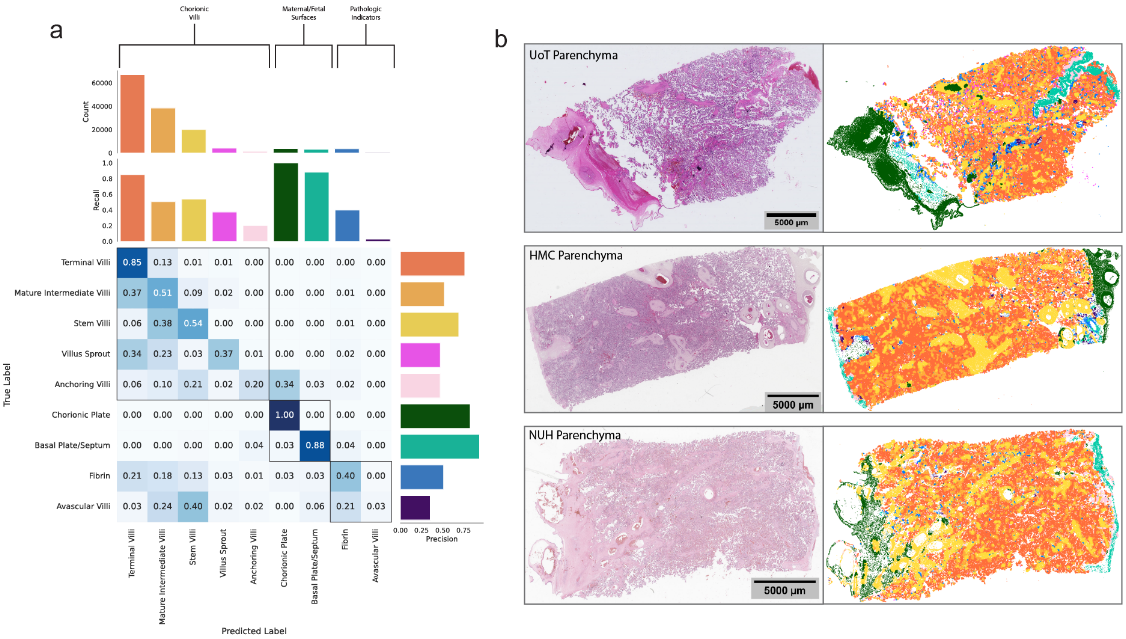
forces the model to understand tissue microstructures in terms of their cellular communities. At inference across a WSI using a laptop CPU, the tissue classification stage classifies ~4,500 nodes per second.

A key benefit of the node classification approach is the freedom for the model to assign different tissue types to different sections of the same continuous structure. Placental tissues grow from one another in a tree-like morphology, so any cross-sectional cut of what appears to be a single structure may contain multiple valid classifications. For example, a cross-section of a mature intermediate villus will likely have terminal villi branching from it but may also contain fibrin, resulting in three different classes. Additionally, the distinctions between villus types are not necessarily discrete; a terminal villus is distinguished from a mature intermediate villus by having >50% of its stroma taken up by capillaries and by having vasculosyncytial membranes (Table 4.1) [61, 81, 86]. However, a section of a mature intermediate villus with unusually many capillaries but no vasculosyncytial membranes, for example, might have nodes that are (mis)classified by the model as a terminal villus section but are not necessarily incorrect.

4.5 Model Evaluation

4.5.1 Test Data Performance

The graph neural network tissue classification model, evaluated across 140,641 cell graph nodes for 9 tissue types, achieves an overall accuracy of 68.56% and a top-2 and top-3 accuracy of 90.53% and 96.13%, with a 0.8859 macro-averaged ROC AUC. Tissue misclassifications (Figure 4.7) primarily fall within developmentally similar microstructures. Misclassifications of villus types are typically confused with other villus types, which correspond to similarities in villus growth and branching morphology [61, 86]. For example, mature intermediate villi, from which terminal villi grow, are mistaken for terminal villi 37% of the time. Likewise, anchoring villi, a subcategory of stem villi with the same cellular composition, are mislabelled as stem villi 21% of the time. Avascular villi, commonly associated with the presence of fibrin [77, 197, 247], are confounded with fibrin 21% of the time. Given the high top-2 accuracy, model misclassifications likely correspond to noise inherent in the biological domain. As previously mentioned, many tissue microstructure types are not discrete categories but sit on a biological continuum. A cell that appears at the blurred boundary of two villus types will have been manually labelled into one category, but the model is not necessarily incorrect in classifying it as the other.



	Terminal Villi	Mature Intermediate Villi	Stem Villi	Villus Sprout	Anchoring Villi	Chorionic Plate	Basal Plate/Septum	Fibrin	Avascular Villi
Terminal Villi	0.85	0.13	0.01	0.01	0.00	0.00	0.00	0.00	0.00
Mature Intermediate Villi	0.37	0.51	0.09	0.02	0.00	0.00	0.00	0.01	0.00
Stem Villi	0.06	0.38	0.54	0.00	0.00	0.00	0.00	0.01	0.00
Villus Sprout	0.34	0.23	0.03	0.37	0.01	0.00	0.00	0.02	0.00
Anchoring Villi	0.06	0.10	0.21	0.02	0.20	0.34	0.03	0.02	0.00
Chorionic Plate	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Basal Plate/Septum	0.00	0.00	0.00	0.00	0.04	0.03	0.88	0.04	0.00
Fibrin	0.21	0.18	0.13	0.03	0.01	0.03	0.03	0.40	0.00
Avascular Villi	0.03	0.24	0.40	0.02	0.02	0.00	0.06	0.21	0.03

Figure 4.7: Tissue node classifier performance on unseen test data. (a) Confusion matrix and precision and recall values of tissue node classifier predictions. Tissue microstructure types are clustered into categories (chorionic villi, maternal/fetal surfaces, and pathologic indicators) and ordered by their test set counts. Categories are highlighted by the topmost brackets and squares within the confusion matrix. (b) Example healthy term placenta WSIs and tissue microstructure predictions on nuclei coordinates (nodes) from each institute.

4.5.2 Generalisability and Standardisation

As mentioned in the previous chapter, generalising models to slides from different institutes remains challenging, and there has been little exploration of model generalisability for GNNs trained on image-extracted features in histology. Typically, in graph deep learning, models are trained to generalise to graphs of new topologies within the same domain by augmenting the graph topology or reducing oversmoothing [248, 249]. In [250], we show how training a GNN using features from a feature extractor which does not generalise well results in poor generalisability of the GNN trained on those features. We show how using graph normalisation to standardise the node features before training and inference is a simple and effective preprocessing step for improving generalisability. In the paper, we present an image-based feature augmentation method that further improves GNN generalisability. However, we also show that GNNs will generalise well to unseen data when the feature extractor performs well on that data.

To assess the generalisability of my GNN methodology, I label 37,149 nodes into tissue microstructures and construct cell graphs across two slides from an institute (NUH) kept unseen to the GNN (Table 4.4). I train a GNN as previously described with and without standardisation to compare the effect on model generalisability. The GNN is only trained on data from the other two institutes, but the cell classifier is trained on data from all three. This isolates the generalisation performance of the GNN when trained on features from a robust feature extractor.

Table 4.4: Graph characteristics of the two intersection cell graphs from a held-out institute. As before, characteristics are calculated after converting edges to undirected and adding self-loops. Homophily metrics are calculated using node homophily [246], with *Homophily* referring to tissue microstructure labels and *Cell Homophily* referring to the cell type of the cell embedding node features.

	Nodes	**Edges**	**Labels**	**Homophily**	**Cell Homophily**	**Degree**
Graph 1	822,959	4,230,387	17,260	0.9687	0.6477	5.1405
Graph 2	1,003,595	5,211,241	19,889	0.9990	0.6397	5.1926

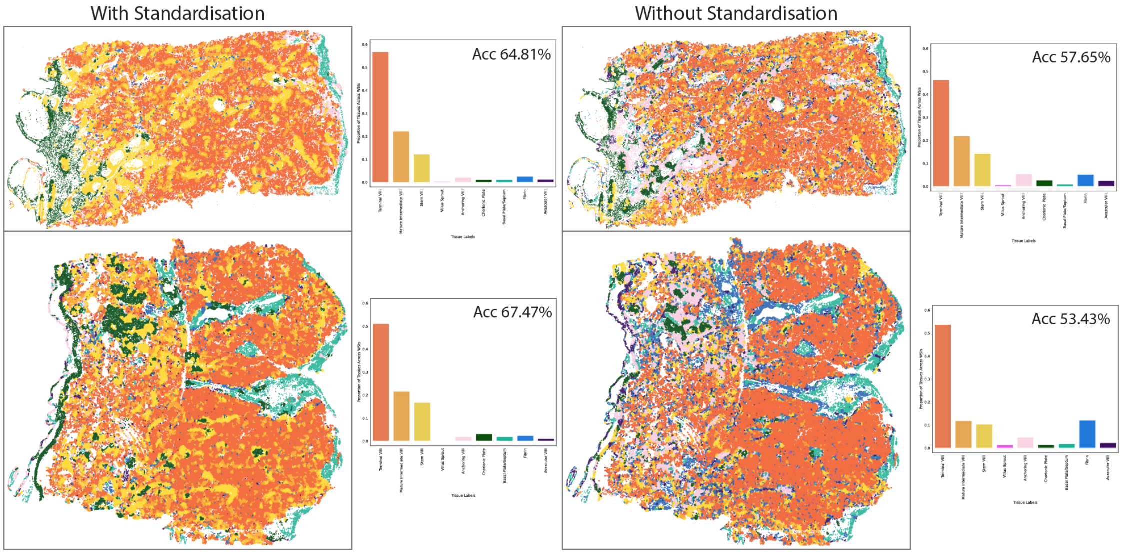


Figure 4.8: Effect of standardisation on tissue predictions and proportions. Training with standardisation improves generalisability and results in accuracies comparable to those on test data from a seen institute. Tissue microstructure predictions are shown across two slides from an institute unseen to the GNN (NUH).

When training and evaluating with standardised node features, the GNN generalises well to the two new slides, achieving top-1 accuracies of 64.81% and 67.47% (Figure 4.8), which are comparable to results on test data. In contrast, the accuracy for the GNN trained without standardised node features drops to 57.65% and 53.43% with more confused predictions for the minority classes. Standardising node features using graph normalisation is an effective preprocessing step for training a GNN to generalise to unseen data.

4.5.3 Hyperparameter Exploration and Ablation

In deep learning, models highly sensitive to changes in their hyperparameters have usually overfit to those hyperparameters and are less likely to perform well on future data. In the following experiments, I change one model or graph construction hyperparameter to assess the model's sensitivity to these changes. For each deviation from the original hyperparameter set, we would expect to see a small variation in performance for most of these experiments. In each case, I compare the best model test accuracy and/or accuracy curves during training to a model trained with hyperparameters described in the methodology. Models are validated every 100 epochs during training using the entirety of the graph rather than subsampling, so for a fair comparison, accuracy curves are shown for both validation and training datasets assessed in this way.

Random node dataset splits. For some node classification benchmark datasets in graph learning, nodes are randomly split into train/validation/test sets across the graph [251]. I train a GNN on the cell graphs where nodes are randomly split into 70/15/15% train/validation/test sets. We can see from the accuracy curves during training (Figure 4.9) that, for cell graphs, splitting nodes randomly results in information leakage. The validation accuracy is the same as the training inference accuracy at every validation step, suggesting that the model is using information from the training set to predict the validation nodes. This result highlights the importance of having a coordinate-based region of the slide designated to dataset splits.

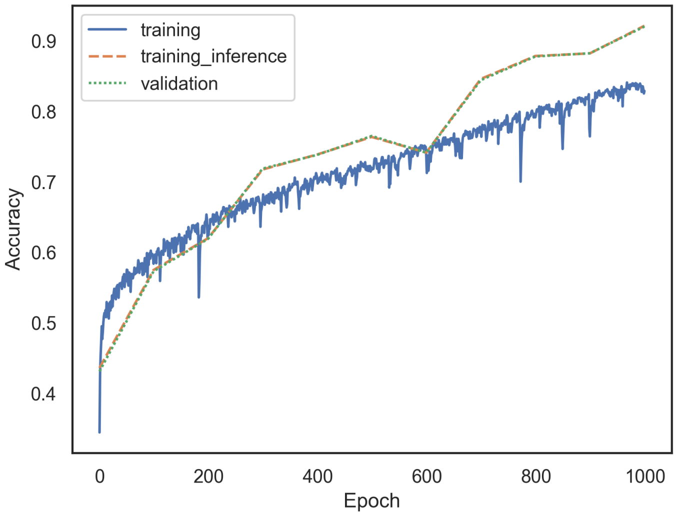


Figure 4.9: Effect of random node splits on accuracy curves. The training and validation inference accuracies are identical at each evaluation epoch, suggesting that there has been information leakage across datasets. A model trained like this will reach maximum validation accuracy by overfitting to the training set and will not generalise well to future data.

The Value of k. Models are trained on cell graphs with edges built from the intersection of a k-NN graph and Delaunay Triangulation. The value of k for the k-NN graph will influence the number of edges in the resulting intersection graph. In Table 4.5, I show for a GNN trained with an intersection k=6 graph, differing values of k at inference can result in better performance. As the model will consider all nodes in a neighbourhood without subsampling during inference, a graph with more edges provides the model with a more expressive neighbourhood. In Table 4.6, I show the performance of a GNN when trained with intersection graphs with different values of k, evaluated at inference with the same value or the previously best-performing value (k=8).

Edge construction. The edge-building algorithm will result in a different graph topology with different graph characteristics. In Table 4.7, I show that an edge construction using Delaunay Triangulation or the intersection method results in the best performance. However, all three methods are comparable.

Table 4.5: GNN performance when trained with an intersection graph with k=6 but differing values of k at inference. Best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

k Inference	Accuracy	Accuracy (Top-2)	ROC AUC
4	65.32%	88.69%	0.8691
6	68.26%	90.35%	0.8828
7	68.60%	90.54%	0.8835
8	**68.77%**	**90.61%**	**0.9042**
9	68.56%	90.53%	0.8859
10	68.43%	90.46%	0.8857
11	68.33%	90.42%	0.8844

Table 4.6: GNN performance when trained with an intersection graph with differing values of k and the same or k=8 at inference. Best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

k Training	k Inference	Accuracy	Accuracy (Top-2)	ROC AUC
4	**4**	65.32%	89.25%	0.8640
4	**6**	67.62%	90.79%	0.8732
6	**6**	68.26%	90.35%	0.8828
6	**8**	68.77%	90.61%	**0.9042**
8	**8**	68.15%	90.52%	0.8872
10	**10**	**69.46%**	**91.29%**	0.8959
10	**8**	68.91%	91.15%	0.8934

Table 4.7: GNN performance when trained and evaluated with graphs where edges were built using a k-NN algorithm, a Delaunay Triangulation or the intersection of the two. Best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

Edge Construction	Accuracy	Accuracy (Top-2)	ROC AUC
k-NN (k=6)	68.44%	90.92%	0.8843
Delaunay	**69.64%**	**90.94%**	0.8950
Intersection	68.77%	68.77%	**0.9042**

Node dropout. At every epoch, a percentage of nodes can be randomly dropped from the graph to provide a regularising effect. In Table 4.8, we see that around a 20% node dropout is optimal for training on histology cell graphs. The accuracy curves in Figure 4.10 show the regularisation effect that node dropout provides, with no dropout resulting in faster overfitting to training data. With too much node dropout the model struggles to learn the training data.

Table 4.8: GNN performance when trained with different levels of node dropout. Around a 20% node dropout is optimal to reduce overfitting and retain enough of a neighbourhood structure for training. The best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

Node Dropout %	Accuracy	Accuracy (Top-2)	ROC AUC
0	67.12%	89.80%	0.8844
20	**68.77%**	90.61%	**0.9042**
40	67.08%	**90.90%**	0.8903
60	63.51%	88.29%	0.8732

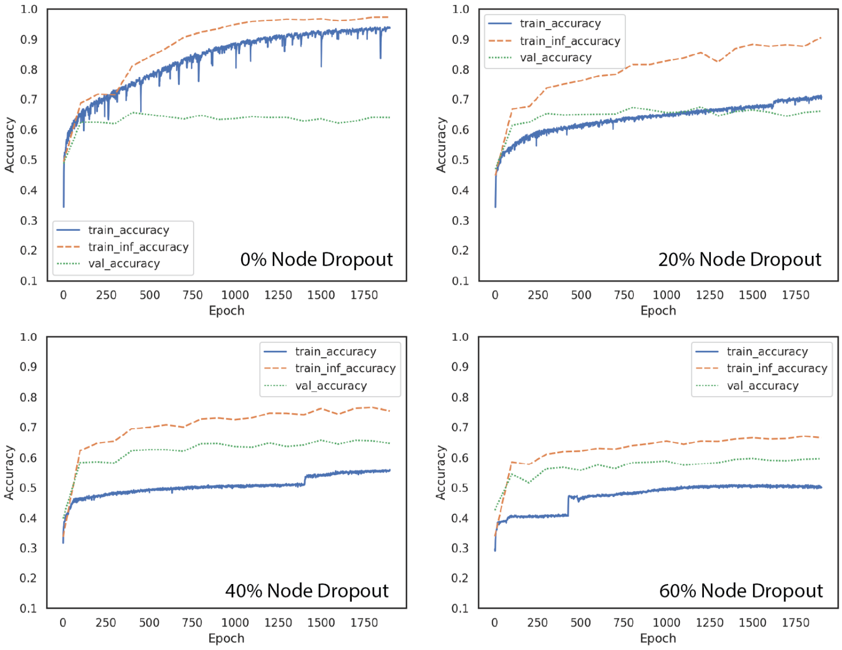


Figure 4.10: Effect of node dropout on accuracy curves. Node dropout provides a regularisation effect and reduces overfitting to training data. However, too much node dropout and the model struggles to train.

Size of receptive field. The receptive field of an inductive GNN is the number of nodes sampled within one neighbourhood across hops for each node prediction. GNN hop size is directly tied to the number of layers in the network, so to increase the receptive field, we need to increase the number of GNN layers. For a ClusterGCN subgraph sampling model, decreasing precomputed subgraph sizes will further restrict this receptive field. In Table 4.9, I show how performance varies across receptive field sizes. Models with larger receptive fields generally perform better but can have prohibitively large GPU memory requirements. For placenta tissue microstructure prediction, a large receptive field is especially important as certain classes can only be distinguished from one another with sufficiently many hops. This is apparent in Figure 4.11, where models with smaller receptive fields confuse the larger stem villus structure (in yellow) with the similar but much larger chorionic plate class (in green).

Table 4.9: GNN performance when trained with different receptive field sizes. A model with a larger receptive field generally performs better but has larger computational requirements. The best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

Receptive Field	Layers	Accuracy	Accuracy (Top-2)	ROC AUC
100 subgraph	4	63.38%	87.52%	0.8709
200 subgraph	8	63.07%	87.63%	0.8744
400 subgraph	16	68.77%	**90.61%**	**0.9042**
800 subgraph	32	**69.05%**	90.52%	0.8888

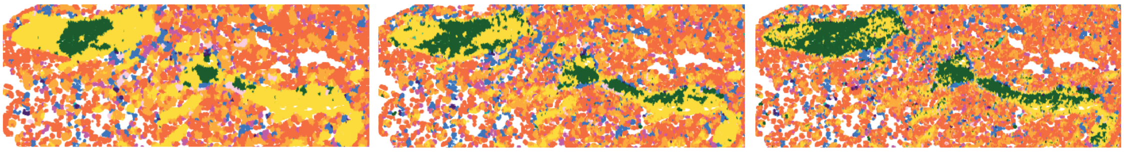


Figure 4.11: Tissue microstructure prediction of models with decreasing receptive fields. Model predictions are more homogeneous within tissue boundaries with larger receptive fields (left) and do not confuse the stem villus structure (yellow) with the much larger chorionic plate class (green) as much as models with smaller receptive fields.

Number of hidden units. The number of hidden units in each layer of the GNN will affect its expressivity. In Table 4.10 and Figure 4.12, I show how a model with too few hidden units cannot learn the data but how a model with too many hidden units will overfit to training data more quickly.

Table 4.10: GNN performance when trained with a different number of hidden units per layer. The best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

Hidden Units	Accuracy	Accuracy (Top-2)	ROC AUC
128	68.32%	90.63%	0.8815
<u>**256**</u>	<u>**68.77%**</u>	<u>90.61%</u>	<u>**0.9042**</u>
512	68.48%	**90.98%**	0.8804

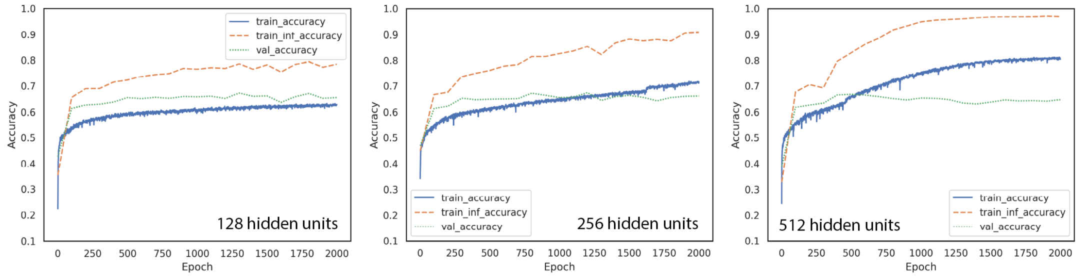


Figure 4.12: Effect of hidden units on accuracy curves. With too few hidden units, the model does not have sufficient parameters to learn the training data. With too many parameters, it overfits to the training data.

Cell class as node features. In all analyses so far, GNNs have been trained on node features extracted from the 64-dimension embedding layer of the cell classifier. To show that these embeddings encode additional useful background and morphological information, in Table 4.11 I compare the performance to a GNN trained on one hot encoded cell classes.

Standardisation. I have shown previously that standardisation is important for training GNNs that generalise well to new cell graph data. In Table 4.12, I show that model performance is comparable on same-institute test data between models trained with and without standardisation. However, for the improvements to generalisability, a model trained with standardisation is preferred.

Table 4.11: GNN performance when trained with node feature cell embedding vectors and one hot encoded cell types. The best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

Node Features	Accuracy	Accuracy (Top-2)	ROC AUC
<u>**Cell Embeddings**</u>	<u>**68.77%**</u>	<u>**90.61%**</u>	<u>**0.9042**</u>
Cell Class	66.04%	89.57%	0.8545

Table 4.12: GNN performance when trained and evaluated with and without graph normalisation standardisation on test data. The best performance in each metric is highlighted in bold, and the hyperparameter described in the methodology is underlined.

Normalisation	Accuracy	Accuracy (Top-2)	ROC AUC
<u>**Standardise**</u>	<u>68.77%</u>	<u>90.61%</u>	<u>**0.9042**</u>
No Standardise	**69.16%**	**91.46%**	0.8950

4.6 A New Node Classification Dataset for Graph Learning in Histology

The release of open-source benchmark datasets has often facilitated the creation of new GNN architectures. From the field's earliest days, the open datasets *Cora*, *Citeseer*, *Pubmed* and *NELL* [252, 253] drove the development of Graph Convolutional Network (GCN) [118], *Reddit* was introduced alongside GraphSAGE [121], and Graph Attention Network (GAT) [254] achieved state-of-the-art on the protein-protein interaction (*PPI*) dataset [255]. However, it was soon apparent that these datasets were not sufficiently complex to challenge GNNs [125, 256]. The datasets had small graphs or a small number of graphs (the largest dataset, Reddit, had 232,965 nodes), they had potentially misleading train, validation, and test splits [251], and they were so well suited to graph learning that randomly initialised, untrained models could perform nearly as well as trained models [257]. In recent years, there has been an influx of graph benchmark datasets, such as the Open Graph Benchmark [125], which address these limitations. They are larger and cover more domains, have realistic train, validation and test splits, and are challenging enough to have not been 'solved' by current GNN architectures [123, 125, 258].

I propose that cell graphs in histology offer a uniquely challenging perspective and are a good new benchmark for developing GNN architectures for several reasons. Firstly, a single cell graph built across one WSI contains over 1 million nodes, node features are defined from an imbalanced distribution of cell class embeddings, and tissue labels also have a high class imbalance, ranging from 0.21% of the data to 40.0%. Secondly, unlike most graph datasets, node clusters are locally homophilic but spatially distributed throughout the graph (globally heterophilic), requiring models to understand same-type node clusters at different locations in the graph. Thirdly, as a WSI is only a small sample of the larger structure with arbitrary cut-offs in edges, out-of-plane cells, and optical and mechanical artefacts, data incompleteness is an inherent feature of this data domain. Finally, the tree-like branching morphology of placental tissue structures will result in separate biological instances which can contain multiple tissue classifications and these structures will vary in size (30-5,000 μm^2). Models must discriminate between small changes across small regions within node clusters but have a large enough receptive field to distinguish between the largest tissues.

To showcase the challenging features of placenta histology cell graphs, I train seven GNN architectures using five scalable graph sampling techniques. Only models compatible with graph sampling can be applied to histology cell graphs, as graphs are too large to fit into GPU memory. During training, for GraphSAGE (with mean aggregation) [121], ClusterGCN [123], GraphSAINT (with random walk sampling) [124], ShaDow [259] and SIGN [260], node embeddings are sampled or precomputed according to original implementations. For models introduced without a sampling technique (GAT [254] and GATv2 [261], I use ClusterGCN subgraph sampling (referred to hereafter as ClusterGAT and ClusterGATv2). During inference, all neighbourhood nodes are used for aggregation. I compare against two baselines, a multi-layer perceptron (MLP) on node embeddings and a randomly initialised GraphSAGE.

Training hyperparameters and graph construction are performed as described above, with a few key differences. For GraphSage, ClusterGAT, and ClusterGATv2, 16-layer models could not fit into GPU memory even with reduced sampling and batch sizes, so fewer layers are used. The cell classification model that generated the node feature cell embeddings is not trained on data from NUH and is trained with a smaller set of augmentations. Graph models are trained without node dropout or graph normalisation, and inference is performed on intersection graphs with k=6. For these reasons, results will differ from those presented so far, but performance is comparable between models.

Table 4.13 shows the test mean accuracies and standard deviations of models over 5 random weight initializations. The randomly initialised baseline performs equivalent to random chance (ROC AUC 0.5), and GNN models outperform the MLP baseline. ClusterGAT and ClusterGATv2 are unstable when trained with more layers, which might contribute to their lower performance score. All scalable GNN architectures: GraphSAGE, ClusterGCN, GraphSAINT, ShaDow, and SIGN are within 2% mean accuracy of each other, with no models surpassing 65% accuracy. These results show that cell graphs can provide a good testbed for developing new GNN architectures fundamentally different from existing scalable methods. I release two cell graphs with manual annotations for the graph learning community [196].

Table 4.13: Performance of scalable GNN architectures. Mean model performance and standard deviations are calculated over 5 random weight initialisations. Bold indicates the best-performing model in each metric.

	Accuracy	**Accuracy (Top-2)**	**ROC AUC**
Random Baseline	5.56±6.24	14.88±11.35	0.459±0.057
MLP Baseline	47.98±0.79	75.22±0.92	0.750±0.003
GraphSAGE-mean	**64.88±0.43**	**88.94±0.38**	0.883±0.005
ClusterGCN	64.24±1.21	88.26±0.82	0.882±0.006
GraphSAINT-rw	63.94±0.23	87.86±0.15	**0.895±0.002**
SIGN	64.77±0.43	88.32±0.42	0.886±0.002
ShaDow	63.04±0.77	86.88±0.74	0.863±0.008
ClusterGAT	58.07±0.61	83.43±0.96	0.851±0.002
ClusterGATv2	57.07±0.65	83.21±0.55	0.854±0.005

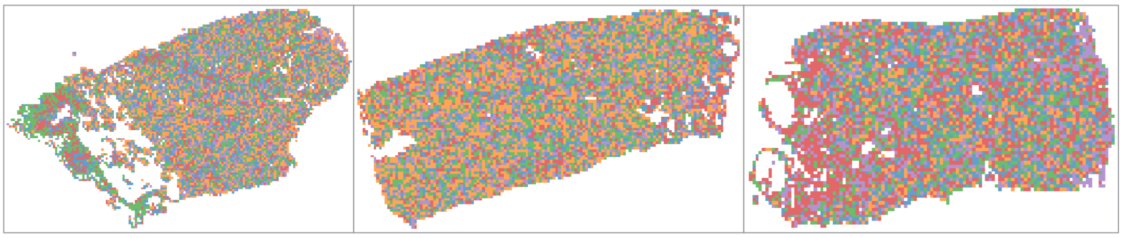
4.7 An Unsupervised Cellular Community Clustering Benchmark

The work presented so far has involved supervised hierarchical prediction of tissue microstructures from communities of cells using graph neural networks. [205] developed a simple, unsupervised method to hierarchically classify image patches into tissue microstructures using cellular communities. They divide a WSI into 200x200 µm patches and use a nuclei localisation model to predict cell coordinates and a cell classification model [262] to classify cells into one of four cell types. They construct a graph across these nuclei coordinates within each patch using an associated Delaunay triangulation [245] and record the proportion of the 10 possible undirected cell-to-cell connections within the patch. They pass this connection information across patches as features to a k-medoids [263] clustering algorithm, which clusters them into 6 clusters. K-medoids is an iterative clustering algorithm similar to k-means [264] except rather than calculating the means of objects in a cluster as the centre, cluster centres (medoids) are chosen as existing data points. This makes it less sensitive to outliers, and medoid features can meaningfully represent clusters. Clusters are analysed for their tissue phenotypes, and labels are assigned manually.

I replicate this approach across placenta slides. As before, I construct a cell graph across the whole slide and record the proportion of cell types within 200x200 µm patches and the 66 different possible cell-to-cell connections from my 11 cell types. Patches with fewer than 10 cells are excluded. I fit two k-medoid algorithms on each slide, one to the proportion of cell types within a patch and another to the proportion of cell-to-cell connections, each predicting 5 clusters. I additionally fit a k-means algorithm to the same data for a comparison. I manually inspect the distribution of patch clusters across each slide to assess performance and use the mean proportion of cell types across clusters to assign tissue phenotypes.

When using the frequency of cell-to-cell connections as features, k-medoids and k-means do not meaningfully cluster patches by tissue phenotypes (Figure 4.13). It may be that the 66 cell-to-cell connections are too complex a feature set by which to cluster the data. However, when using the cell type proportions, the patches broadly separate into macro tissue structures (Figure 4.14). From manual inspection, clusters separate into distal villi (villus sprouts, terminal villi and mature intermediate villi), stem villi and the chorionic plate, fetal vessels, and the basal plate. The clusters assigned by k-means have more distinct cellular profiles and fit into better tissue phenotypes than those assigned by k-medoids. Unsupervised clustering of patches using cell graphs is an effective and simple method to identify tissue structures across a slide. A downside of this approach is the requirement for manual cluster and patch size selection and manual assignment of clusters. However, it can be a good benchmark for other supervised or unsupervised methods. This exploration would additionally benefit from a quantitative comparison to ground truth tissue microstructure annotations and GNN predictions by mapping them to patches.

K-medoids on cell to cell connections



K-means on cell to cell connections

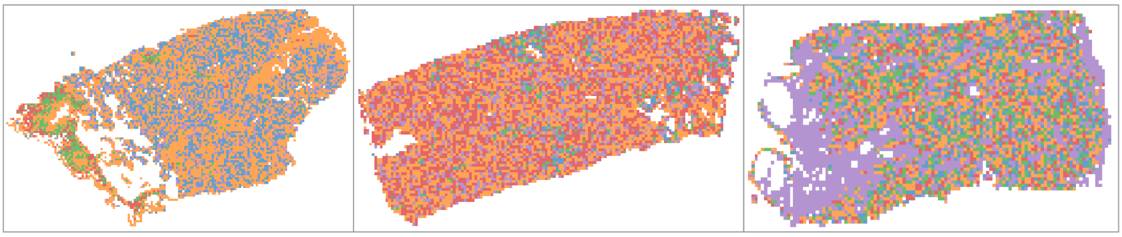


Figure 4.13: Patches clustered by k-medoid and k-means clustering using the 66 cell-to-cell connections in each patch. Neither clustering algorithm is able to cluster the patches using cell connection features into distinct tissue phenotypes.

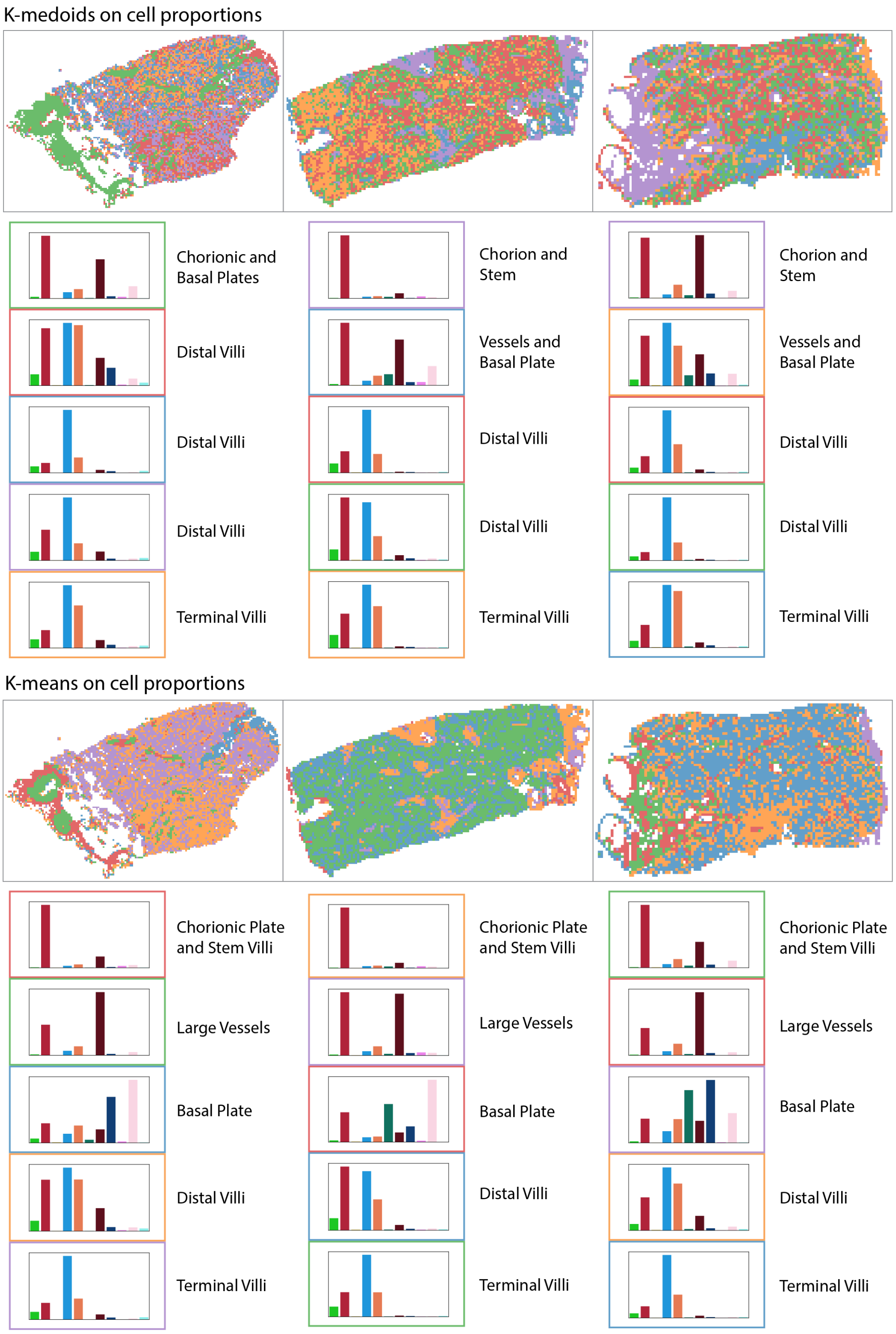


Figure 4.14: Patches clustered by k-medoid and k-means clustering using the 11 cell type proportions in each patch. Both clustering algorithms broadly separate the patches into macro tissue phenotypes. The mean cellular proportions of each cluster are shown below each slide, with their borders coloured by their corresponding cluster class on the slide.

4.8 Conclusion

Building upon the last chapter's nuclei coordinates and cell class outputs, I have presented a cell-graph-based hierarchical method for classifying tissue microstructures in parenchyma histology of the term placenta. I use a scalable message-passing GNN on cell graphs constructed across whole slides to aggregate cellular communities into one of 9 tissue microstructure types. I show how the model generalises to slides from an unseen institute and highlight how node feature standardisation contributes to training a robust model. I present a comprehensive hyperparameter exploration to show that the model is not highly sensitive to hyperparameter and graph construction and to showcase the challenges GNNs have on cell graph data. Further, I compare the performance of seven scaleable state-of-the-art GNNs and describe a cell graph dataset release to the graph learning community for developing new scalable GNN architectures. Finally, I explore how an unsupervised hierarchical tissue phenotyping method over patches is a simple and effective benchmark for identifying larger tissue regions.

In the context of the literature, this is the first method for automated classification and localisation of tissue microstructures in placenta histology and is one of few methods using GNNs for node classification on cell graphs built across entire whole slide images. Together with the cellular phenotypes, these tissue microstructure phenotypes can be used for rich, biologically explainable slide-level prediction.

4.9 Limitations

The hierarchical nature of this cell-to-tissue graph approach allows us to model and validate tissue predictions as the composition of their cellular communities. However, it also couples the tissue model's performance and outputs to the performance and outputs of the cell model. Any changes or improvements to the cell parts of the pipeline will require retraining of the tissue model. Whilst tissue inference is computationally very fast, taking just a few minutes to predict across a whole slide on a CPU, it requires running the nuclei localisation and cell classification models that are computationally more intensive. Further, poor cell model performance, such as when stain differences cause a domain shift, will result in cellular embeddings with lower informational quality as inputs to the GNN. To some extent, this can be overcome by training the GNN with regularisation, such as graph normalisation and node feature augmentation as shown in MEGA [250], but ideally, we would look to improve the cell model. Finally, as this is a node classification task on nuclei, there will only be tissue predictions for regions with nuclei. In practice, however, nuclei are almost always present in the tissue regions we would want to classify, including lesions.

Further limitations include those specific to using GNNs for node classification on histology cell graphs. The current scalable GNN architectures struggle with labels which are locally homophilic but distributed throughout the graph (globally heterophilic), and their fixed receptive field sizes make predicting both very small and very large tissue structures a challenge.

As with the cell models, this analysis could be extended to predict on placenta slides from earlier gestational ages by training with annotations for immature intermediate villi and mesenchymal villi and for the other placental slide types. Unsupervised hierarchical cell graph-based tissue phenotyping could also be further explored to avoid the need for manual annotation.

You must have chaos within you
to give birth to a dancing star.

— Friedrich Nietzsche

5

Quantifying Healthy Variation in the Term Placenta

Contents

5.1 Introduction

There is a lack of normal baselines for understanding healthy variation within the term placenta [23, 265]. The placenta is a notoriously heterogeneous organ both at different sampling sites within the same organ and across placentas from uncomplicated pregnancies [16, 23, 51]. The genetic mosaicism inherent to the placenta may go some ways towards explaining this heterogeneity [20]. The placenta has redundancy and a reserve capacity that allows it to tolerate a degree of biological insult, and it exhibits adaptive responses in suboptimal conditions [13, 43]. However, the thresholds for these tolerances are not currently clear. There have been efforts in recent years to establish clinical consensus for these thresholds [103, 266], but further large-scale quantitative analyses are needed [265]. Historically, the placenta has been an understudied and poorly understood organ despite its importance [4, 13, 31, 42–45, 55, 267].

In this chapter, I aim to present new, biologically rich quantitative metrics for assessing placental health and understanding healthy variability in the term placenta. I use the previous chapters' cellular and tissue microstructure prediction models to quantify variability across 30 WSIs of healthy term placentas from three institutes. I validate model predictions against the agreement scores of four practising perinatal pathologists, against expectations from our current understanding of placental biology and reference ranges in placental literature.

In the following sections, I present the expected cellular and tissue microstructure biology of the normal term placenta, including any reference ranges, and I summarise the existing literature for healthy placental phenotyping. I describe the data used in this analysis and restate the methodology. I show how tissue model prediction and ground truth annotations follow similar agreement patterns to the agreement between perinatal pathologists and how cellular and tissue predictions, including the cellular composition of tissues, match biological expectations and reference ranges. Finally, I use UMAP to show how the internal representation of cell and tissue microstructure classes mirrors biological class similarity.[1]

---

[1] Sections titled *Methodology*, *Comparison to Perinatal Pathologists* and *Predictions Match Expectations from Placental Biology and Literature* have been adapted and expanded on from [175] (Nature Communications 2024). Sections titled *Related Work*, *Expected Cell and Tissue Microstructures*, *Model Representation Reflects Biological Similarity* and *Limitations* are new work.

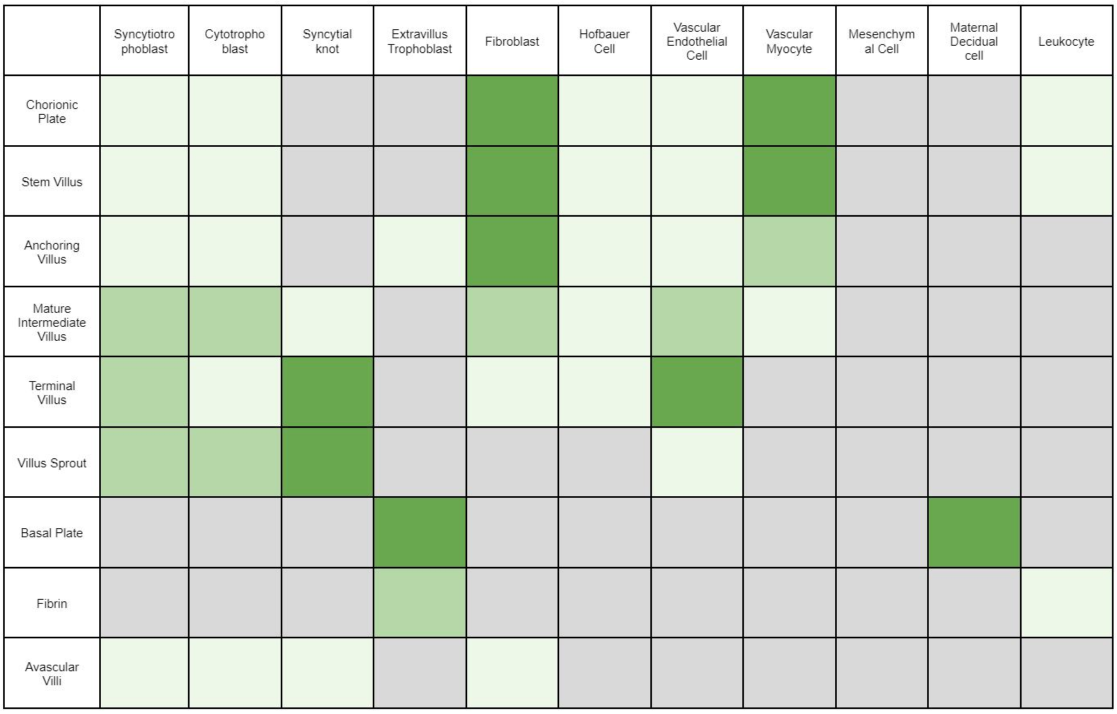
5.2 Expected Cell and Tissue Microstructures in Healthy Term Placentas

For tissue microstructures, there are reference ranges for the proportional surface area of the chorionic villus types and thresholds for clinical significance of the included indicators of pathology (Table 5.1) [56, 64, 81, 197, 198]. Ranges indicate that most of the placenta surface area on a slide of a healthy term placenta consists of terminal villi, followed by mature intermediate villi, stem villi, and a small number of villus sprouts. Healthy term placentas should have less than 2.5% of villi as avascular villi, and less than 25% of the slide area should be fibrin [197, 198]. No ranges are reported for the proportion of the chorionic plate, basal plate, and anchoring villi, as these will vary depending on slice and organ preparation.

Table 5.1: The expected surface area of tissue microstructures across a healthy term placenta slide. Ranges are shown alongside corresponding literature references.

Tissue Microstructure	Expected Surface Area
Terminal Villus	30-60% [56, 81]
Mature Intermediate Villus	17-32% [56, 64, 81]
Stem Villus	9-25% [56, 64, 81]
Villus Sprout	0-1% [56]
Fibrin	<25% [197]
Avascular Villus	<2.5% [198]

For the cellular composition of tissue microstructures, there are no reference ranges, but we may use the morphological and functional definitions reported in the literature [56, 61, 86, 268, 269]. In Figure 5.1, I show for each of the tissue microstructures included in this analysis, whether the cell type is not present in that tissue, whether the cell type is proportionally low in that tissue, proportionally medium in that tissue, or proportionally high in that tissue, relative to other tissue microstructures. Note that this categorical representation is only intended as a rough guide based on tissue microstructure schematics and descriptions and should not be used to definitively classify those structures. In summary, the smaller, more vascular chorionic villi, such as terminal villi, will contain more trophoblasts and vascular endothelial cells due to their higher surface area and capillary density. The larger, more structural chorionic villi, such as stem villi, will contain more structural cell types, such as fibroblasts and vascular myocytes, in addition to leukocytes in the fetal blood of fetal vessels. Avascular villi will have the same composition as their original villus type but without vascular cells. Extravillus trophoblasts generally populate fibrin, and the basal plate will contain all maternal decidual cells and some extravillus trophoblasts.



	Syncytiotrophoblast	Cytotrophoblast	Syncytial knot	Extravillus Trophoblast	Fibroblast	Hofbauer Cell	Vascular Endothelial Cell	Vascular Myocyte	Mesenchymal Cell	Maternal Decidual cell	Leukocyte
Chorionic Plate	░	░	☐	☐	█	░	░	█	☐	☐	░
Stem Villus	░	░	☐	☐	█	░	░	█	☐	☐	░
Anchoring Villus	░	░	☐	░	█	░	░	▒	☐	☐	☐
Mature Intermediate Villus	▒	▒	░	☐	▒	░	▒	░	☐	☐	☐
Terminal Villus	▒	░	█	☐	░	░	█	☐	☐	☐	☐
Villus Sprout	▒	▒	█	☐	☐	☐	░	☐	☐	☐	☐
Basal Plate	☐	☐	☐	█	☐	☐	☐	☐	☐	█	☐
Fibrin	☐	☐	☐	▒	☐	☐	☐	☐	☐	☐	░
Avascular Villi	░	░	░	☐	░	☐	☐	☐	☐	☐	☐

Figure 5.1: Cell type composition of tissue microstructures in WSIs of healthy term placentas. Categorical cell type is approximated from schematics and descriptions in placenta literature. Grey indicates that the cell type is not present in that tissue, light green indicates that the cell type is proportionally low, medium green indicates that it is proportionally medium, and dark green indicates that it is proportionally high, all relative to other tissues.

As there are upwards of a million cells in a term placenta histology sample, it has been infeasible to manually develop reference ranges for the expected proportion of cell types. However, we can approximate them from the reference ranges for the expected proportional area of tissue microstructures and their cellular composition. Due to the high proportion of terminal villi, we may expect to find a high proportion of syncytiotrophoblast nuclei and vascular endothelial cells, and the low expected proportion of avascular villi further evidences this. Due to the maturity of term samples, we would not expect to find many cytotrophoblasts or undifferentiated mesenchymal cells [61, 86]. As these are healthy samples, there should not be many leukocytes to suggest an inflammatory response [270, 271]. The proportion of extravillus trophoblasts and maternal decidual cells is affected by the amount of basal plate in the sample, which will vary by slice and organ preparation. However, an unusually high proportion of extravillus trophoblasts may indicate an excess of fibrin [197].

5.3 Related Work

Identification of cellular and tissue microstructure elements for slide-level phenotyping in placenta histology is commonly performed using stereology [46]. Stereology is an unbiased sampling method for generalising from micro-anatomical observations in small regions to the 2D slide or 3D organ level. Regions are selected by randomised, systematic sampling across multiple slides and sites, and features of interest are manually annotated within these regions. In the placenta these features often include structure type (i.e. terminal villi, syncytial knots, fetal capillaries), volume densities, vasculosyncytial barrier thickness, and counts [46–48] and allow for estimates of the organ's diffusive capabilities and fetoplacental angiogenesis from villus growth and maturity [46]. A downside of stereology is the need for expert, labour-intensive manual annotation within sampled regions, meaning that studies are usually limited to smaller sample sizes and there is an assumption that sampled regions are representative. However, one could argue that these same downsides exist with the manually annotated training data used to train supervised deep learning models. There may be a potential collaboration between stereology and deep learning (a kind of automated stereology) where annotated regions for stereology are used as training data and stereology estimation techniques are used to generalise from model predictions to the 3D organ-level structure.

The only deep learning method applied to healthy slide-level prediction in placental histology is GestAltNet for predicting the gestational age of a slide [51].

Here, authors use a VGG19 model with weighted attention on 16 patches from each slide to predict the slide-level gestational age. They compare against a baseline VGG19 model trained to predict gestational age directly on patches. In both cases, the median gestational across patch sets or individual patches is used as the slide-level prediction. All training, validation, and test data are preselected to contain regions of terminal villi. Their method achieves a mean absolute error (MAE) of 1.08 weeks, better than the 1.45 of the baseline model on test data. They compare the historical patterns of gestational age diagnosis by eight pathologists across 9.5 years and note a variability of around 1.37 weeks, assuming an even random distribution of placentas among pathologists. However, it is unclear if these are perinatal pathologists or general surgical pathologists for which interobserver reliability is lower [29].

5.4 Whole Slide Image Data

I use 30 parenchyma slides from the placentas of 30 singleton healthy term pregnancies from all three institutes for whole slide healthy cell and tissue microstructure quantification. Given the heterogeneity and resiliency of the placenta, I select slides from and define ‘healthy’ as term placentas from pregnancies with no adverse health outcomes during or after pregnancy and where pathology reports and second assessment by perinatal pathologist state that histological sections of the parenchyma were ‘normal’ and that ‘villi correspond to gestational age’. As an example, a portion of these healthy placentas from HMC were from pregnancies which had suspected placenta accreta from a 1st-trimester ultrasound, thereby qualifying them for submission to microscopic examination, but with no resulting complications (Table 5.2).

Table 5.2: Patient characteristics for each placenta. Patient characteristics and relevant clinical data for each placenta defined as healthy and used in this analysis. *Clinical data and exact gestational age for slides from NUH are undisclosed as per the data transfer agreement.

Institute	Gestational Age	Relevant Clinical Data
UoT	42+0	Placenta vallate, hypocoiled cord
HMC	40+0	Suspected placenta accreta
HMC	36+5	Suspected placenta accreta
HMC	40+1	Suspected placenta accreta
HMC	41+5	Suspected placenta accreta
NUH x 25	37 – 42 (term)*	Histologically normal

5.5 Methodology

I run the trained nuclei localisation, cell classification, and tissue node classification models of the prior chapters across the 30 healthy term WSIs. To summarise (Figure 5.2), this involves sectioning a WSI into overlapping 1600x1200 (177.44x133.08 µm) pixel images, using a RetinaNet model to identify the nuclei coordinates in each image, extracting 200x200 (22.18x22.18 µm) pixel images around each nucleus and using a ResNet-50 model to classify the image into one of 11 cell types. From the nuclei coordinates and cell predictions, a cell graph is constructed across the whole slide and input into a ClusterGCN model to classify each node's tissue microstructure. For each slide, this results in a set of nuclei coordinates where each nucleus has an associated cell type prediction and 64 dimension cell embeddings and a tissue microstructure prediction and 64 dimension tissue embeddings. These predictions are then used to quantify healthy variation across the 30 WSIs.

Whilst I have applied the methodology described so far to placenta histology, the codebase has been intentionally written to be organ agnostic with guides for data annotation, entry script tutorials, and demos on a small sample slide. This codebase, HAPPY (for Histology Analysis Pipeline.PY), is being used internally by other members of the lab for liver, adipose and kidney histology with some exploratory multi-organ analysis. I hope that this project continues to mature into a high-utility toolset for the group.

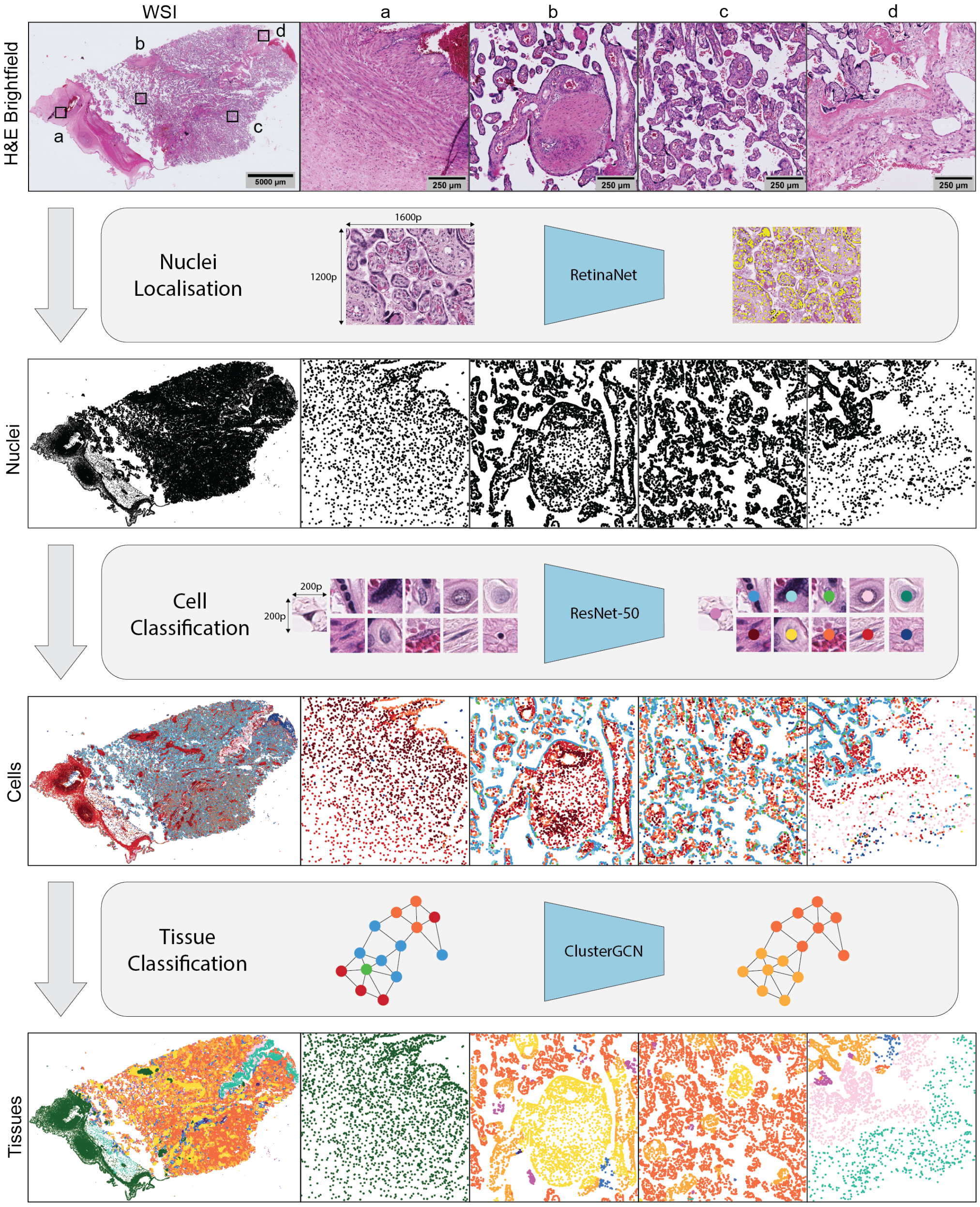


Figure 5.2: HAPPY workflow. A hematoxylin and eosin (H&E) stained whole slide image (WSI) is first sectioned into overlapping 1600x1200 (177.44x133.08 µm) pixel images and passed to an object detection RetinaNet model, which identifies the nuclei in these images. 200x200 (22.18x22.18 µm) pixel images centred on each nucleus are classified into one of 11 cell types by a ResNet-50 model. The 64-dimension embeddings from the cell classifier and their nucleus coordinates are used to build a cell graph across the whole slide image. The cell graph is input into a ClusterGCN graph neural network, which classifies the tissue microstructure to which each cell belongs. Images a-d show characteristic tissue regions of the WSI: (a) chorionic plate, (b) stem and distal villi, (c) distal villi, (d) basal plate and anchoring villi.

5.6 Comparison to Perinatal Pathologists

Four practising expert perinatal pathologists were asked to perform a similar labelling task across tissue microstructures to assess the accuracy of my ground truth tissue microstructure training annotations and to judge the relative difficulty of identifying placental tissue types. Each pathologist was shown a series of images containing a centred tissue microstructure out of 180 total images with the task of labelling the tissue type of that centred structure. Images were generated from a random, class-balanced subset of my ground truth annotations, and pathologists were blind to the original annotation and each other's labels. Images were cropped to display contextual background, similar to the context a 16-layer message-passing GNN may see, but it was expected that tissue types were identifiable from their cellular composition alone.

Participants were first presented with a Standard Operating Procedure (SOP) and a tutorial. These documents detailed the labelling setup, what data would be collected, and how it would be used, and included links to current literature relevant to placental tissue microstructures. Participants were invited to a project in the browser-based software LabelBox [272] where they were sequentially presented with images and could choose one of 12 tissue types for that image (Figure 5.3). Alternatively, they could state that the type was unclear or not listed, and they could leave a comment. Participants were informed that all images came from a healthy term placenta. After completion, their tissue type labels were compared for Cohen's kappa [273] agreement scores against each other, the original annotations, and the model's prediction.

Across all tissue microstructure types, pathologists have a moderate mean agreement score of 0.55 kappa, with low-moderate kappa values for mature intermediate villi (0.468), villus sprouts (0.371), avascular villi (0.154), and anchoring villi (0.051). Pathologists disagree with their majority-voted label (Figure 5.4) at least 50% of the time for anchoring villi and avascular villi, highlighting the difficulty of identifying these structures.

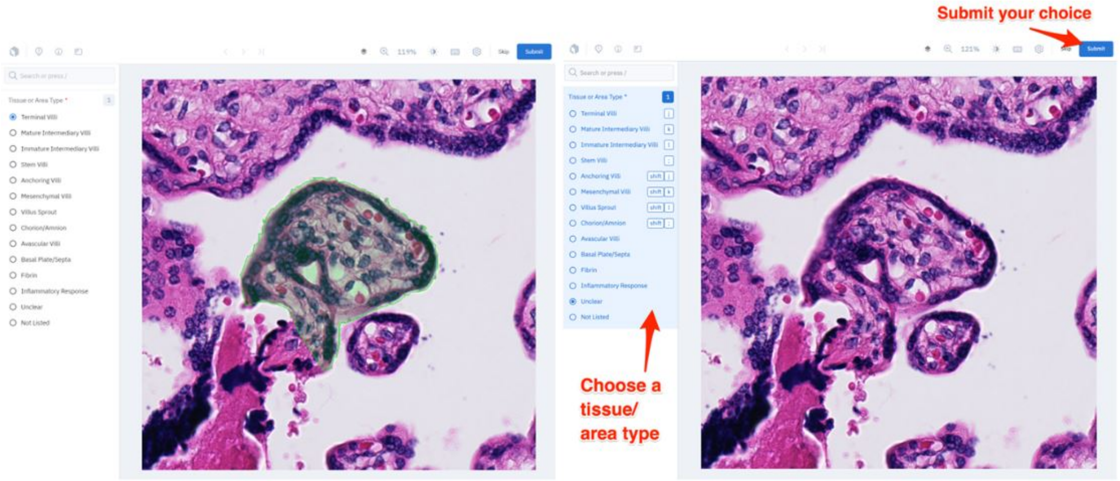
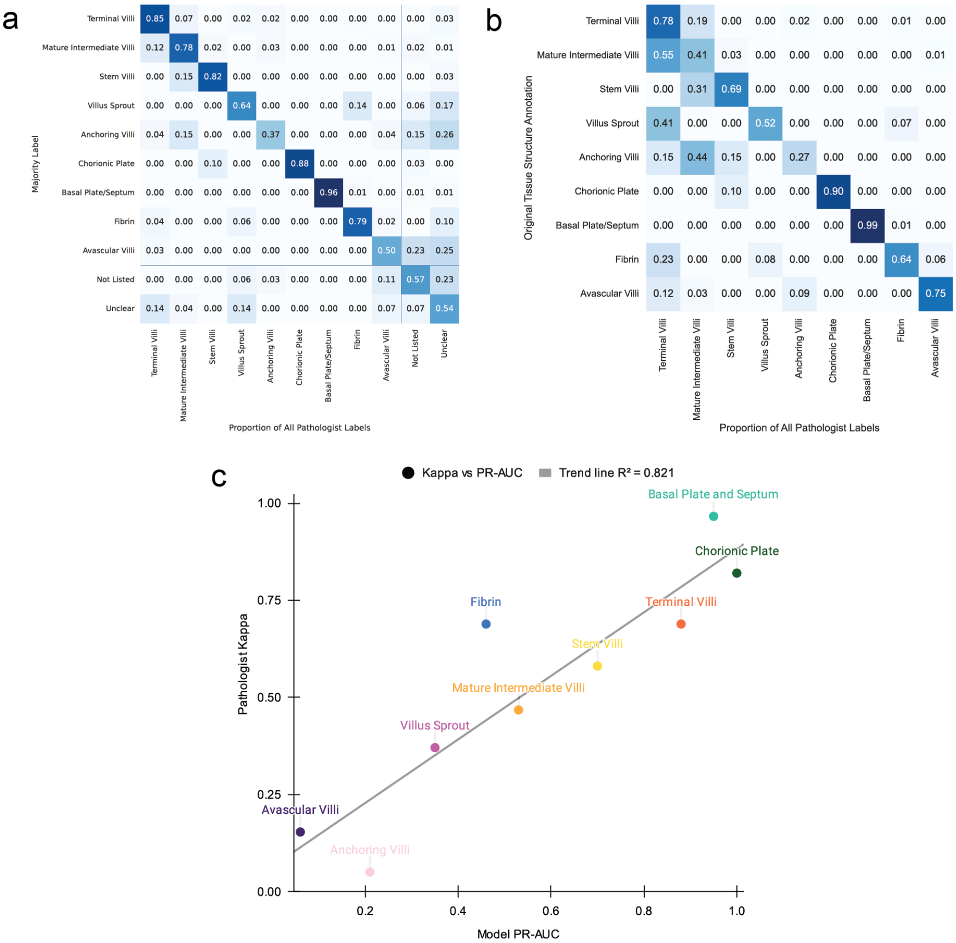


Figure 5.3: User interface for pathologist tissue microstructure labelling task. For each of the 180 images shown to pathologists, the tissue microstructure to label was centred and highlighted with a segmentation overlay, which could be toggled on and off. Pathologists were given a multiple choice selection of labels, could add a comment, and submit their answer.

Taking the pathologists' majority label as the gold standard, there is a 0.61 kappa value to the ground truth annotations, indicating a slightly better agreement than the inter-pathologists' agreement. For 7/9 tissue microstructures, the pathologists' labels match the annotations >50% of the time (Figure 5.4). There is a strong match for terminal villi (78%), avascular villi (80%), chorionic plate tissue (90%) and basal plate tissue (99%). Of the two structures with <50% label match, mature intermediate villi (41%) and anchoring villi (27%) were among the structures with the lowest inter-pathologist agreement, as described above.

Finally, I contrast the tissue model's PR-AUC (Precision Recall Area Under Curve) values against mean pathologist agreement scores for each tissue microstructure type (Figure 5.4). The PR-AUC values have a strong positive correlation (R^2=0.821) with the mean Cohen's kappa score between the pathologists, suggesting the model's predictions are on par with perinatal pathologists for this task. Additionally, pathologists label disagreement (Figure 5.4) show similar patterns to model confusion.

Pathologist disagreement in this task is not unexpected as specific, structure-by-structure tissue classification is not part of pathology investigations, partly because this is not feasible for humans at scale. Nonetheless, the gestalt or organised whole, i.e. tissue type and morphology assessment in aggregate, is a key part of pathology reporting and disease prediction. Specific tissue classification performance comparable to human experts shows models can accurately quantify placenta biology in a way that is likely relevant to pathology detection. These results highlight the potential of large-scale deep learning methods to identify abnormalities in placental microstructures that are too subtle to be recognised by routine light microscopy examination.



	Terminal Villi	Mature Intermediate Villi	Stem Villi	Villus Sprout	Anchoring Villi	Chorionic Plate	Basal Plate/Septum	Fibrin	Avascular Villi	Not Listed	Unclear
Terminal Villi	0.85	0.07	0.00	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.03
Mature Intermediate Villi	0.12	0.78	0.02	0.00	0.03	0.00	0.00	0.00	0.01	0.02	0.01
Stem Villi	0.00	0.15	0.82	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
Villus Sprout	0.00	0.00	0.00	0.64	0.00	0.00	0.00	0.14	0.00	0.06	0.17
Anchoring Villi	0.04	0.15	0.00	0.00	0.37	0.00	0.00	0.00	0.04	0.15	0.26
Chorionic Plate	0.00	0.00	0.10	0.00	0.00	0.88	0.00	0.00	0.00	0.03	0.00
Basal Plate/Septum	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.01	0.00	0.01	0.01
Fibrin	0.04	0.00	0.00	0.06	0.00	0.00	0.00	0.79	0.02	0.00	0.10
Avascular Villi	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.23	0.25
Not Listed	0.00	0.00	0.00	0.06	0.03	0.00	0.00	0.00	0.11	0.57	0.23
Unclear	0.14	0.04	0.00	0.14	0.00	0.00	0.00	0.00	0.07	0.07	0.54

	Terminal Villi	Mature Intermediate Villi	Stem Villi	Villus Sprout	Anchoring Villi	Chorionic Plate	Basal Plate/Septum	Fibrin	Avascular Villi
Terminal Villi	0.78	0.19	0.00	0.00	0.02	0.00	0.00	0.01	0.00
Mature Intermediate Villi	0.55	0.41	0.03	0.00	0.00	0.00	0.00	0.00	0.01
Stem Villi	0.00	0.31	0.69	0.00	0.00	0.00	0.00	0.00	0.00
Villus Sprout	0.41	0.00	0.00	0.52	0.00	0.00	0.00	0.07	0.00
Anchoring Villi	0.15	0.44	0.15	0.00	0.27	0.00	0.00	0.00	0.00
Chorionic Plate	0.00	0.00	0.10	0.00	0.00	0.90	0.00	0.00	0.00
Basal Plate/Septum	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00
Fibrin	0.23	0.00	0.00	0.08	0.00	0.00	0.00	0.64	0.06
Avascular Villi	0.12	0.03	0.00	0.00	0.09	0.00	0.00	0.00	0.75

Figure 5.4: Pathologist agreement and model confusion across tissue types. (a) Inter-pathologist confusion matrix showing for each image with a majority pathologist label (y-axis) how much variance there was against that majority (x-axis). Lines indicate the additional options for pathologists to say the tissue was unclear or not listed. (b) Confusion matrix which shows the proportion of matching pathologists' labels with ground truth annotations. (c) Model performance Precision Recall Area Under Curve (PR-AUC) scores plotted against mean pathologist Kappa values by tissue type showing a strong positive correlation ($R^2 = 0.821$).

5.7 Predictions Match Expectations from Placental Biology and Literature

To further validate HAPPY, I compare cell and tissue microstructure whole slide predictions to expectations from our current understanding of placental biology and against any reference metrics reported in the literature. In Figure 5.5, I show the variability of predicted cells as a proportion of all cells across a WSI for the 30 parenchyma WSIs of healthy term placentas. These predictions reflect the expected internal anatomy of a healthy term placenta. The high proportion of syncytiotrophoblast nuclei (>40%) relative to villus stromal cells matches the expected large surface area to volume ratio of an effective villus tree system optimised for diffusive exchange [86]. The <1% proportion of undifferentiated mesenchymal cells and low proportion of cytotrophoblasts (4%-13%) are characteristic of the late maturation stage of the placenta samples [56, 61, 81, 86]. The low proportion of leukocytes (2%-4%) are below clinical thresholds for pathological relevance [266, 270].

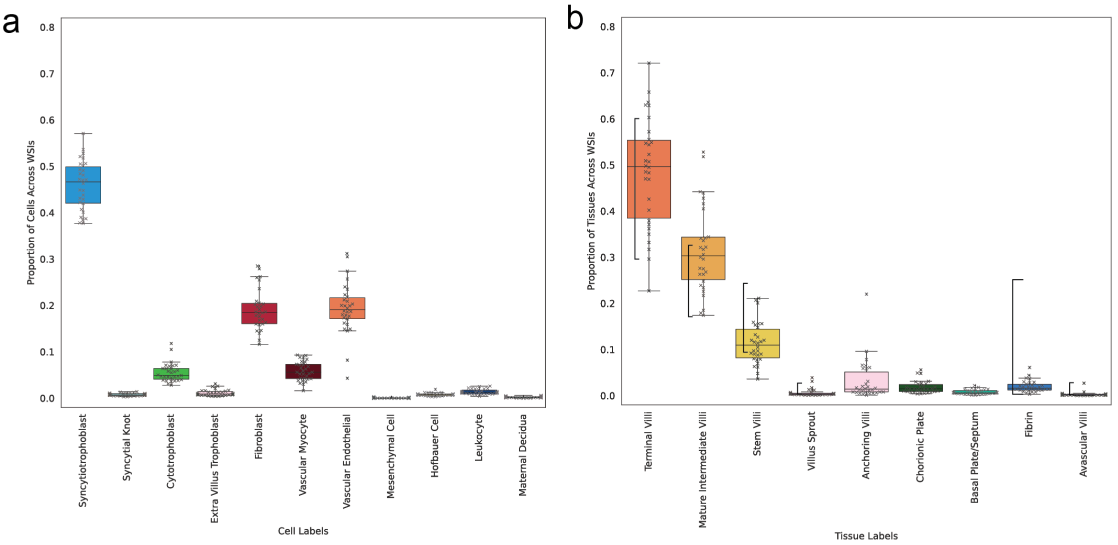


Figure 5.5: Predicted cell and tissue microstructure type proportions across 30 whole slide images (WSI) from healthy term placentas. (a) Cell proportions across WSIs. (b) Tissue microstructure proportions across WSIs with vertical braces showing the expected ranges reported in the literature for healthy term placentas. Box centre line represents the median and whiskers are drawn up to 1.5 times the interquartile range. Each WSI datapoint is shown by a cross marker.

Next, I analyse the variability of predicted tissue microstructures as a proportion of all tissues across a WSI for the 30 parenchyma WSIs of healthy term placentas. The proportion of villus microstructures predominantly fall within the ranges reported in the literature for healthy term placentas (Figure 5.5) [8, 81, 86, 198]. In most slides, terminal villi are the most common villus type, comprising an expected 30-60% of the slide's tissue structures. Likewise, mature intermediate villi and stem villi are the 2nd and 3rd most common tissue structures, predominantly falling within the expected ranges of 17-32% and 9-25%, respectively. Excessive regions of fibrin or avascular villi are indicators of pathologic processes [197, 266] and the proportions of fibrin and avascular villi are $<25\%$ and $<2.5\%$, respectively, which are below the clinical thresholds for pathological relevance [197, 199, 266].

The hierarchical cell-to-tissue structure of HAPPY allows us to inspect tissue predictions in terms of their underlying cellular communities. As the four chorionic villus types and the chorionic plate are defined and distinguished from one another by their cellular compositions, we can use this cell-to-tissue structure to validate predictions in a biologically interpretable way. In Figure 5.6, I show the mean predicted cellular composition for the four chorionic villus tissues and the chorionic plate across the 30 WSIs of healthy term placentas. The predicted cellular proportions match current descriptions and schematics reported in placenta literature [61, 86]. As villus types develop from one another in a tree-like structure, their cellular proportions shift along the tree in a continuum and model predictions recapitulate this continuum. For example, the terminal villi, which form the tips of the villus tree and are the primary sites for maternal/fetal diffusive exchange, are characterised by their $>50\%$ capillary stromal volume [61, 86]. As such, their cellular composition contains the highest proportion of vascular endothelial cells. Conversely, the stem villi and chorionic plate, which form the trunk and base of the villus tree, respectively, support the villus structure and contain the highest proportion of structural fibroblasts and vascular myocytes. Their larger vessels will contain proportionally more fetal blood with higher proportions of predicted leukocytes corresponding to the increased leukocytes found in fetal blood at later gestations [274]. The mature intermediate villi have predicted cellular proportions between stem and terminal villi, which corresponds to their place in the chorionic tree. Finally, villus sprouts contain the highest proportion of trophoblasts with little to no stromal core cells, as expected [61, 86].

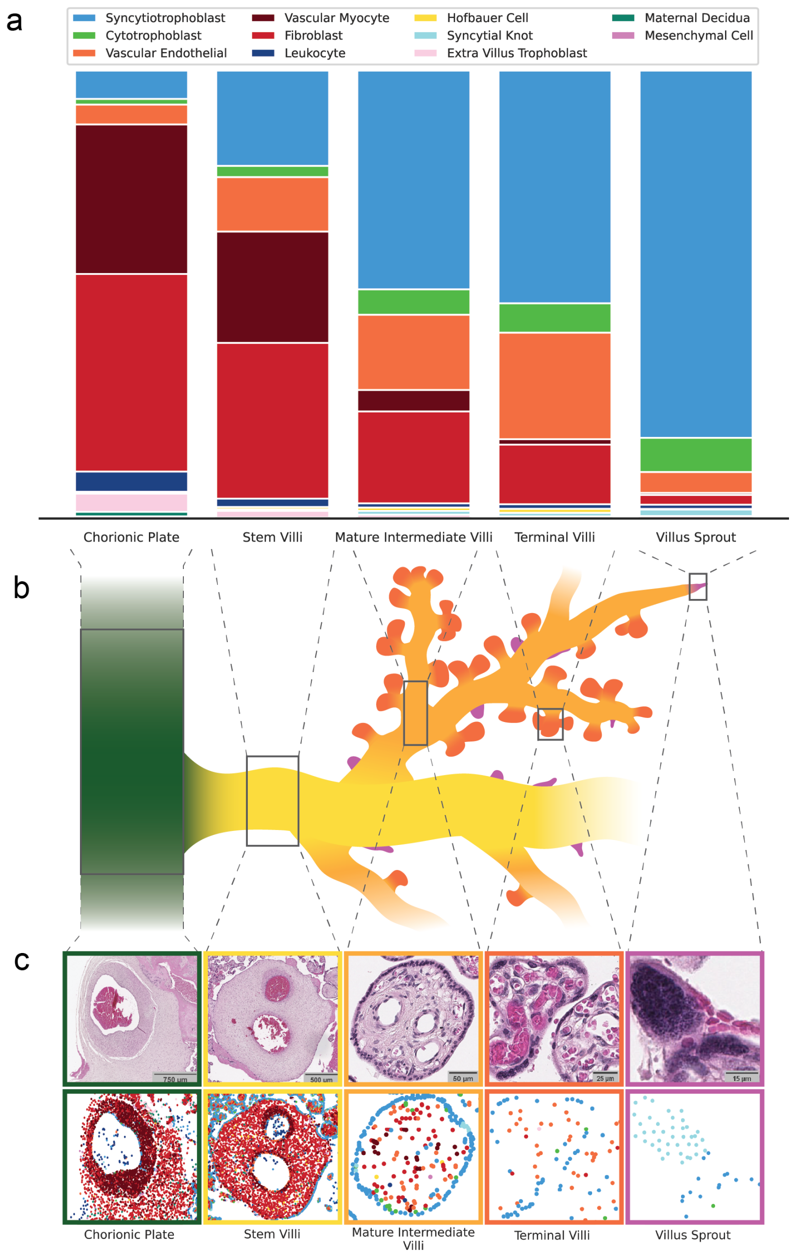


Figure 5.6: Proportions of predicted cell types within predicted chorionic villus tissues and their corresponding locations in a villus tree schematic. (a) The mean proportion of predicted cell types within each predicted term chorionic villus tissue for 30 whole slide images. (b) A villus tree schematic showing how each villus structure relates to and grows from other villus structures. (c) Example histology and cell predictions for each villus structure.

5.8 Model Representations Reflect Biological Similarity

We would like our models to learn an internal representation of classes which mirrors similarity in the underlying biological phenotypes. In part, this can be observed from the previous chapters' top-2 or top-3 accuracies and confusion matrices. Another way to explore this representation is by visualising the latent vector space using a dimensionality technique such as UMAP [275]. If the model has a good internal representation, we expect similar structures to be closer in embedding space. As these are models trained in a supervised way, we would not expect to see very distinct clusters, but model decision boundaries should at least fall across similar classes. The UMAP visualisation additionally allows us to explore how the tissue microstructure prediction model shifts the input vectors (cell embedding vectors) across higher dimensional space towards tissue microstructures. In the following analysis, I fit a UMAP to the 64-dimension embedding vectors of the cell and tissue microstructure prediction models and plot the 2-dimension reduced embeddings coloured by the cell and tissue predictions.

The models should broadly group structures by their biological categories. For the cells, the trophoblasts and villous stromal cells should be grouped. For the tissue microstructures, chorionic villi should follow the tree-like continuum from chorion to stem villi, mature intermediate villi, terminal villi, and villus sprout, with the basal plate and pathologic indicators being distinct.

The cell model's internal representation reflects biological similarity (Figure 5.8). For each slide, the trophoblasts are grouped on one side of the plot, and the villus stromal cells are on the other. The syncytial knots lie next to the syncytiotrophoblast nuclei, which are next to the cytotrophoblasts. The villus stromal cells have a vascular myocyte cluster next to the fibroblasts and vascular endothelial cells. The tissue microstructure predictions on the cell embeddings also reflect the previous section's cellular compositions. Vascular myocytes are predominantly in chorion and stem villi predictions, fibroblasts follow a gradient across villi types, and vascular endothelial cells and trophoblasts are in the terminal villi.

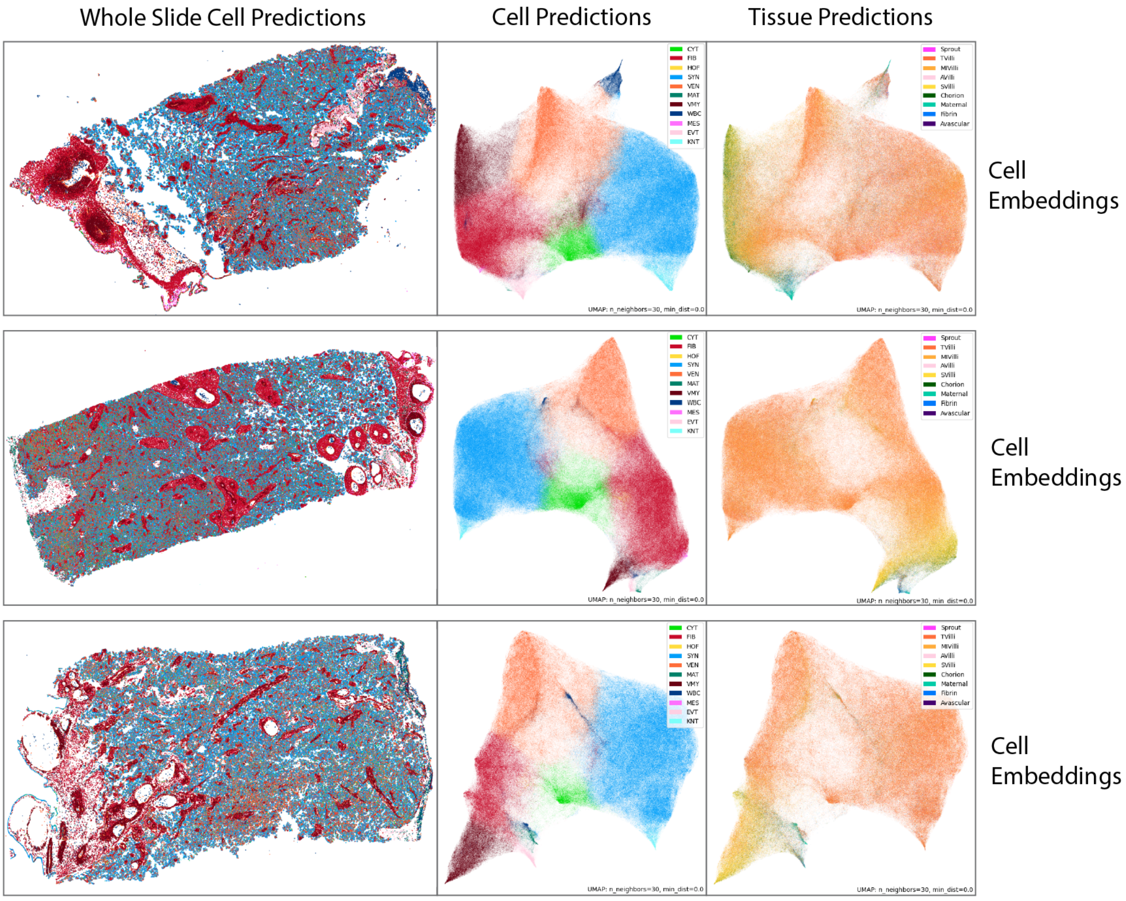


Figure 5.7: Cell predictions and corresponding UMAP plots on one healthy term slide from each institute. A UMAP is fitted on cell embeddings for all nuclei across a slide. Each point is coloured by the predicted cell and tissue microstructure.

A similar pattern is observed in the tissue microstructure model's representation (Figure 5.8). For each slide, the villus continuum from the terminal villus to the stem villus is preserved. The chorionic plate clusters near to stem villi or is distinct, and the basal plate is a separate cluster. Fibrin and avascular villi are offshoots from the villi. As before, the cell predictions on tissue embeddings reflect expected cellular composition. The chorion and stem villi contain fibroblasts and vascular myocytes, the mature intermediate villi have a mix of villus stromal cells and trophoblasts, and the terminal villi contain the syncytial knots and trophoblasts. The tissue microstructure prediction model uses cellular community and composition to identify tissues in a biologically meaningful way.

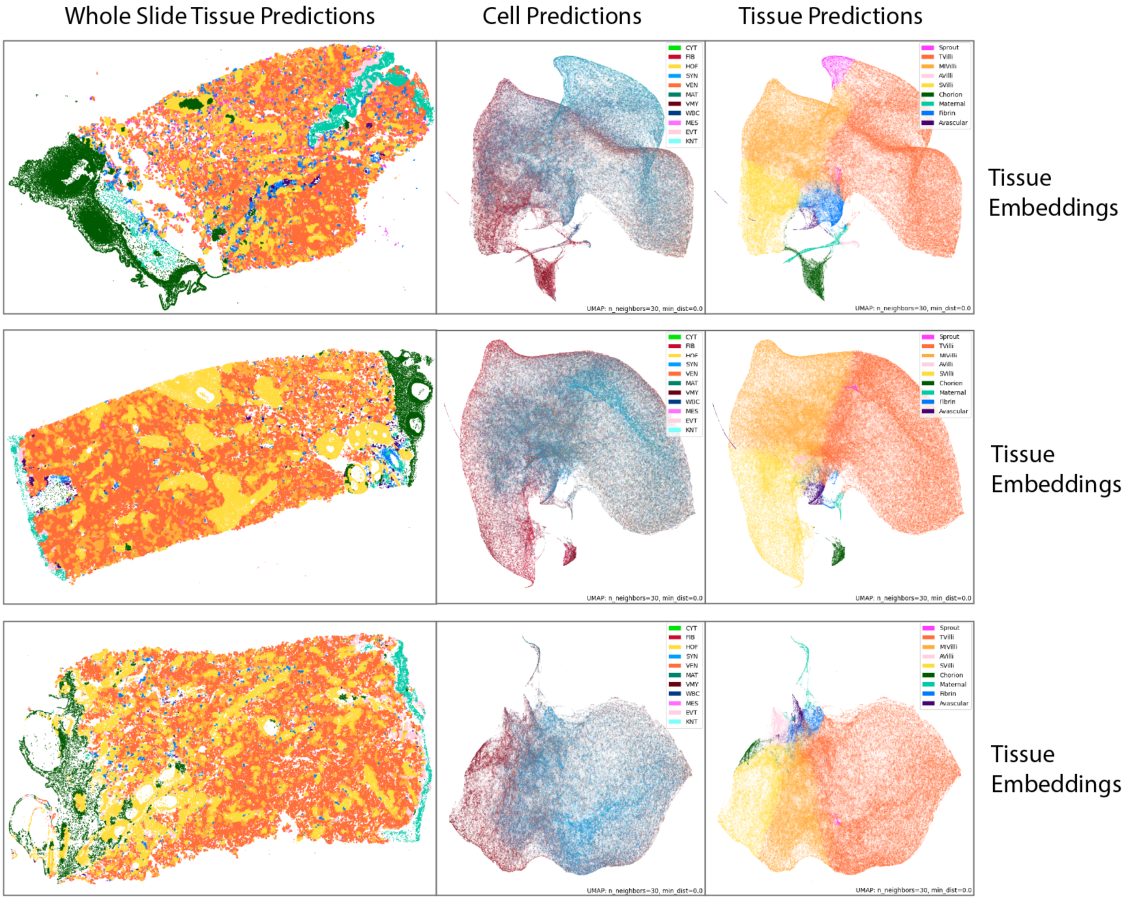


Figure 5.8: Tissue microstructure predictions and corresponding UMAP plots on one healthy term slide from each institute. A UMAP is fitted on tissue embeddings for all nuclei across a slide. Each point is coloured by the predicted cell and tissue microstructure.

5.9 Conclusion

In this chapter, I have explored using aggregated subslide metrics for quantifying healthy variation in the term placenta at the slide level. I ran the cell and tissue microstructure prediction models of the last two chapters across 30 WSIs of healthy term placentas from three institutes. I validated the tissue model's outputs against the agreement scores of four practising perinatal pathologists, showing similar performance and patterns of disagreement. I validated the aggregated whole slide cell and tissue metrics against our expectations from biology and placental literature. The proportion of cells, tissue microstructures, and cellular composition of tissue microstructures matches these expectations for healthy term placentas. I further show how the internal representation of the models reflects the underlying biology.

In the context of the literature, this is the first automated, scalable method for combining cellular and tissue microstructure outputs to quantify placenta histology and provides the first set of whole slide cellular and tissue microstructure benchmarks for healthy term placentas. It is a step in understanding micro-anatomical variation that is tolerated by the placenta and provides a sliding scale of explainable biological phenotypes to measure health and disease.

5.10 Limitations

With over a million cells and tens of thousands of tissue microstructures on a placenta slide, the presented methodology provides rich biological phenotypes to quantify a placenta sample. However, in aggregate, these only provide baseline metrics for the 30 analysed healthy term placentas. Additionally, as slides are drawn from clinical datasets collected as part of routine pathology investigation, there may be unknown changes to their underlying morphologies that differentiate them from healthy controls. Future work would benefit from more healthy term slides from more institutes collected explicitly as healthy controls.

Right now, this analysis only predicts cellular and tissue microstructure densities and proportions, but it could be extended by translating and comparing against the metrics used in placenta stereology. Metrics such as the size and distance between terminal villi, number of vasculosyncytial membranes and diffusion distance, number of knots per terminal villus, size of vessels, etc., would be useful for clinical practice and placenta research. For example, this could be used to investigate the expected villus changes under the two surface plates. In healthy placentas, villi under the chorionic plate are sparser and smaller in diameter, appearing accelerated, and villi under the basal plate are larger with looser stroma appearing immature [23], but this has not been quantified. These metrics may be approximated using existing predicted features. For example, edge distance can be used to judge villus distances, syncytiotrophoblast nuclei to segment individual villi, vascular myocytes and vascular endothelial cells to measure vessel sizes and capillary densities.

Not to laugh, not to lament,
not to detest, but to understand.

— Baruch Spinoza

6

Identifying Parenchymal Lesions

Contents

6.1 Introduction

A large part of placental pathology reporting involves identifying and assessing the severity of lesions on placenta slides [3, 8, 266]. Taken together, lesions across all slides from the same placenta form a pattern of injury, which indicates a pathological process [103, 276]. Placental pathology reporting is essential for clinical management of maternal and newborn health. It informs immediate clinical treatment and predicts recurrence risk in subsequent pregnancies, long-term maternal and child adverse outcomes, and explains underlying causes of pregnancy loss [1–9]. Perivillous fibrin, for example, has a high recurrence risk [2, 197], and maternal vascular malperfusion can predict maternal hypertension and cardiovascular disease a decade after pregnancy [277].

Associating lesions with pathologies and outcomes, however, is challenging as the placenta is highly resilient to injury with a reserve capacity and inbuilt redundancy [13]. Term pregnancies with normal outcomes have been found to contain low-severity lesions up to 35.7% of the time and high-severity lesions 9.1% of the time [21]. Further research is needed to quantify the biological changes associated with these lesions on a large scale and verify thresholds for when lesion severity becomes pathologic and leads to adverse outcomes [8, 54]. Further, placenta pathology reporting is labour-intensive and requires specialised pathologists [29]. Tools assisting placenta pathology reporting could be invaluable for clinical workflows and placenta research.

In this chapter, I explore the use of cellular and tissue microstructure phenotypes to identify parenchymal lesions. Following an introduction to the lesions in this analysis and related work in identifying placental parenchymal lesions, I divide the chapter into two sections. In the first section, I apply the cell and tissue microstructure prediction models from previous chapters across a range of slides containing one or more parenchymal lesions. I first compare predictions on term slides with clinically significant placental infarction to those of healthy term placentas and find significant differences. I then describe the full parenchymal lesion dataset and a parallelised workflow for efficiently running the inference pipeline across this data. Finally, I use the aggregate cell and tissue microstructure predictions across all lesion slides to analyse the phenotypic biological changes caused by these lesions. In the second section, I introduce a novel unsupervised graph compression autoencoder (GCAE) for classifying regions of interest around lesions in a slide. The model includes a custom one hop node compression algorithm for efficiently selecting evenly distributed supernodes across a graph. I validate this methodology for cellular and tissue microstructure embedding reconstruction and downstream region prediction.[1]

6.2 Parenchymal Lesions of Interest

I focus on the four most common parenchymal lesions: infarction, perivillous fibrin, avascular villi, and intervillous thrombosis (Figure 6.1). These lesions can be found alone or in combination on a slide, are associated with specific cellular and tissue microstructure changes and are an important part of placental pathology reporting as markers of maternal vascular malperfusion and fetal vascular malperfusion.

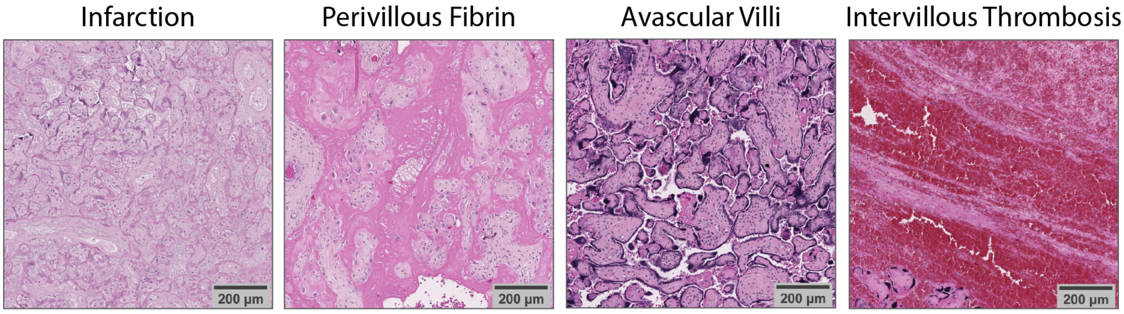


Figure 6.1: Example histology of each placenta parenchymal lesion of interest. Infarction is characterised by its pale villi with nuclear degeneration, perivillous fibrin is characterised by its fibrinoid material surrounding villi, avascular villi are villi without vasculature or stromal content, and an intervillous thrombus is a large bleed in the intervillous space with varying levels of fibrosis.

6.2.1 Infarction

Placental infarction is a lesion of the placental parenchyma whereby a region or regions of villi undergo ischaemic coagulative necrosis [84]. Caused by a

---

[1] Section titled *Pilot Case Study on Placental Infarction* has been adapted and expanded on from [175] (Nature Communications 2024). All other sections in this chapter are new work.

disruption of the maternal circulation within the placental space, it is a key reported finding in pathological investigation as a marker of maternal vascular malperfusion [84, 103, 266, 278]. When clinically significant [266], it is associated with maternal hypertensive disease, abruption, preterm birth, and fetal growth restriction [278–280].

Given the biological changes caused by placental infarction [84], at the cellular level, we would expect to see fewer cells found in healthy distal villi such as syncytiotrophoblast, cytotrophoblast, fibroblast and vascular endothelial cells and more extravillus trophoblasts and leukocytes. In terms of tissue microstructures, we would expect there to be fewer distal villi, such as the terminal and mature intermediate villi and, in their place, there should be a larger proportion of fibrin and villi without vasculature (avascular villi). As the age of infarction will affect the degree of nuclear degeneration (a young infarction will exhibit less nuclear degeneration) [278], these changes will sit along a continuum.

6.2.2 Perivillous Fibrin

Perivillous fibrin deposition is an excessive buildup of fibrinoid material in the intervillous space, causing villi to become sclerotic and necrose over time. It is considered pathologic when more than 25% of the slide is surrounded by fibrinoid material (increased perivillous fibrinoid deposition at 25%-50% and massive perivillous fibrinoid deposition at >50%) [197, 281]. When clinically significant, it is associated with fetal growth restriction, intrauterine fetal demise, and adverse neurologic outcomes [281, 282] with a high recurrence rate in subsequent pregnancy [2, 197].

At the cellular level, we would expect to see an increase in extravillus trophoblasts as these invade the fibrinoid material [197] and, depending on the severity, a decrease in villus stromal cells as villi are cut off from the surrounding maternal blood. As with placental infarction, we would expect fewer distal villi as these necrose and lose their vasculature, as well as an increase in the fibrin type as the fibrinoid material fills the intervillous space.

6.2.3 Avascular Villi

Avascular villi is a lesion of fetal origin where villi become ischemic due to an obstructed blood flow in a vessel at an earlier point in the villus tree or umbilical cord, vascular damage, or an inflammatory pathology [80]. Ischemic villi undergo a process known as villous stromal-vascular karyorrhexis [80, 247] where there is cell death of villus vascular and stromal cells as they lose their internal oxygen supply. The villus is considered avascular at the final stage of this cell death, but trophoblasts remain viable due to their contact with maternal blood. When there are more than three clusters of more than two avascular villi, it is considered clinically significant and is associated with fetal growth restriction, fetal central nervous system injury and stillbirth as a marker of fetal vascular malperfusion [247, 266].

Given the vascular and stromal cell changes in avascular villi, we would expect to find fewer vascular and stromal cells at the cellular level. However, there should be little to no change in trophoblast densities. If an inflammatory pathology causes the lesion, then there may be an increase in leukocytes. At the tissue microstructure level, there should be a reduction in distal villi, but as avascular villi are clinically significant in small quantities, this may not be apparent across a slide.

6.2.4 Intervillous Thrombosis

Intervillous thrombosis, also called haematoma or thrombohematoma, is a coagulation of blood within the intervillous space [283]. It has many possible causes, including haemorrhage from fetal capillaries, abnormal or stasis of blood flow within the intervillous space, or maternal hypercoagulability [283, 284]. It is associated with placenta increta, percreta, preeclampsia and maternal hypertension [284, 285]. Intervillous thrombi are present in a fifth of term placentas but, when alone, do not usually present with an adverse outcome [21, 284].

At the cellular level, we expect to see an increase in leukocytes in the coagulating blood. We would not expect to see large changes to the tissue microstructures when a thrombus is the only lesion on a slide. However, there will be small regions where villi can become avascular or infarcted if they have been pushed aside or entrapped by the thrombus [283].

6.3 Related Work

There are four methods in the literature for predicting placental lesions in histology using machine learning or deep learning. The first is the previously mentioned work on hierarchical decidual vasculopathy prediction [52]. The second is a method for predicting the presence of infarction, perivillous fibrin, intervillous thrombosis or no lesion at the slide level from patches [53]. Here, authors use an ImageNet pretrained EfficientNetV2L to extract patch features and a subnetwork that outputs patch scores for each class. A weighted average of scores across the slide is passed to a four-layer fully connected network to predict the presence of each lesion at the slide level. The third method trains a support vector machine classifier on extracted image features to predict the presence of maternal vascular malperfusion lesions [54]. Image features are extracted using an ImageNet pretrained ResNet18 on non-overlapping patches and are passed to an SVM for binary classification of maternal vascular malperfusion. The slide-level prediction is obtained by majority voting across patches. The fourth method uses a convolutional neural network to predict the severity of distal villous hypoplasia on slide patches, obtaining a slide-level prediction by averaging patch scores [286].

6.4 Section 1: Using Cellular and Tissue Phenotypes

For the first exploratory analysis of deep learning for identifying parenchymal lesions, I use the cellular and tissue microstructure prediction models on slides with one of the four lesions of interest. After an initial pilot study comparing predictions from slides of healthy term placentas to slides from term placentas with clinically significant placental infarction, I apply the models to a larger diseased cohort. This cohort, consisting of 315 parenchyma slides, is derived from placentas of 2nd, 3rd trimester, and term pregnancies where pathology reports include one of the four parenchymal lesions.

6.4.1 Pilot Case Study on Placental Infarction

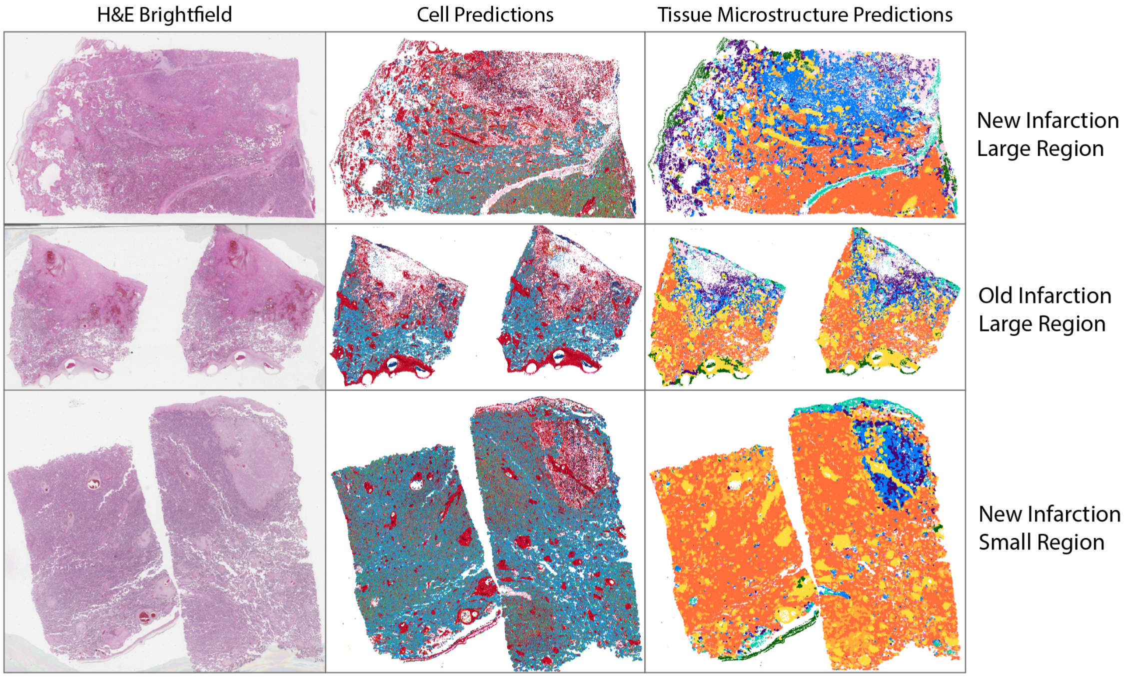


Figure 6.2: Three slides with placenta infarction and corresponding cell and tissue microstructure predictions. Each slide displays different infarction characteristics. The top row contains a newer infraction over a large portion of the slide, the middle row contains an older infarction, as seen by the nuclei sparsity, and the bottom row contains a newer infarction over a small portion of the slide.

I perform an initial pilot case study for slides from term placentas with clinically significant placental infarction to assess whether the aggregated cellular and tissue microstructure predictions would provide useful metrics for assessing placental parenchymal lesions. I use 12 term WSIs from eight singleton term placentas with clinically significant placental infarction in pathology reports at original institutes from UoT and HMC. A secondary assessment of slides is provided by another perinatal pathologist at another institute. These 12 slides are carefully selected to contain both young and old infarctions on the slide, as this will affect the degree of nuclear degeneration, and a range of infarction sizes. See Figure 6.2 for an example of these differences. I test the significance of these cell and tissue microstructure differences independently between the 30 WSIs of healthy term placentas and these 12 WSIs with placenta infarction using two-sided Welch's t-test.

Contrasting the proportion of cells across our samples (Figure 6.3), we see that syncytiotrophoblast (p=0.001), fibroblast (p=0.002), and vascular endothelial cells (p<0.001) are nominally significantly fewer in placentas with infarction and extravillus trophoblast cells (p=0.001) and leukocytes (p=0.002) are significantly higher. In terms of the tissue microstructures, there are fewer terminal villi (p<0.001) and mature intermediate villi (p=0.03) and more fibrin (p=0.002) and avascular villi (p=0.001). Additionally, proportions of fibrin and avascular villi surpass the healthy expected ranges reported in the literature for 2/12 and 8/12 WSIs with placental infarction, respectively. Similarly, 9/12 and 4/12 WSIs with placental infarction have proportions of terminal villi and mature intermediate villi below expected ranges for healthy term placentas.

Given that placental infarction will result in fewer total nuclei across a slide, I additionally compare the number of predicted cell and tissue microstructure counts per mm^2 area of tissue on the slide (Figure 6.4). I estimate this area by splitting the slide into non-overlapping patches and aggregating the area of patches containing at least one nucleus prediction. There is significantly less total nuclei density (p<0.001) in the WSIs with placental infarction. There are nominally significant results for cell and tissue type densities across the slides. After Bonferroni multiple testing correction, syncytiotrophoblast (p=0.002), extravillus trophoblast (p=0.01), vascular endothelial cells (p<0.001), leukocytes (p=0.006) and total density (p=0.003) remain significant. Likewise, for the tissue microstructure types, terminal villi (p=0.002), mature intermediate villi (p=0.009), anchoring villi (p=0.03), fibrin (p=0.007) and avascular villi (p=0.003) remain significant.

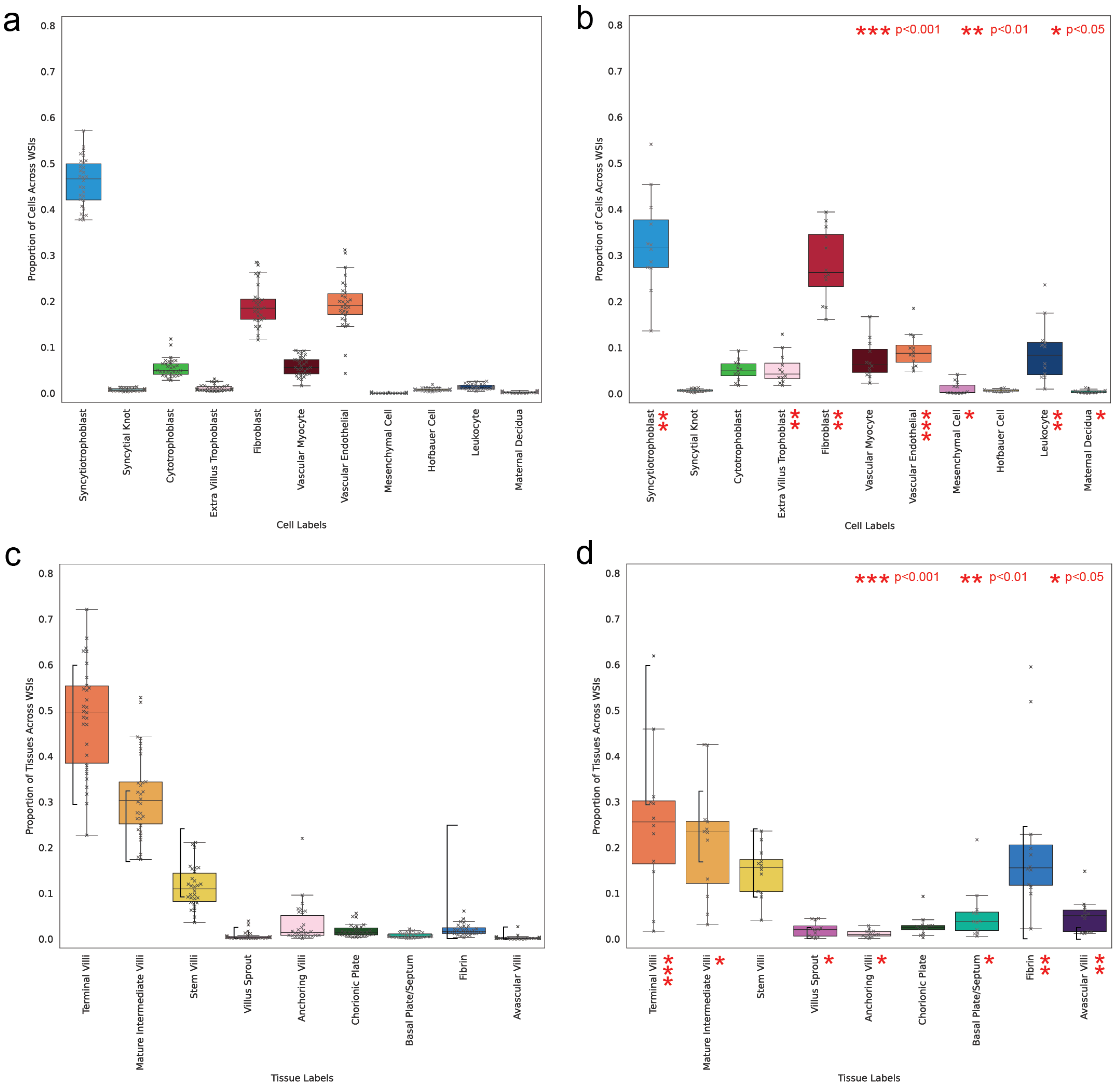


Figure 6.3: Predicted cell and tissue microstructure proportions across 30 whole slide images (WSI) of healthy term placentas and 12 WSIs of placentas with placental infarction. (a) Cell proportions across healthy term placentas, (b) cell proportions across term placentas with placental infarction, (c) tissue proportions across healthy term placentas, (d) tissue proportions across term placentas with placental infarction. The box centre line represents the median, and whiskers are drawn up to 1.5 times the interquartile range. Each WSI datapoint is shown by a cross marker. Expected healthy ranges for tissue microstructures, as reported in the literature, are shown by black vertical bars. Nominal significant differences in cell and tissue structures between the two groups are calculated using two-sided Welch's t-test and shown by red asterisks.

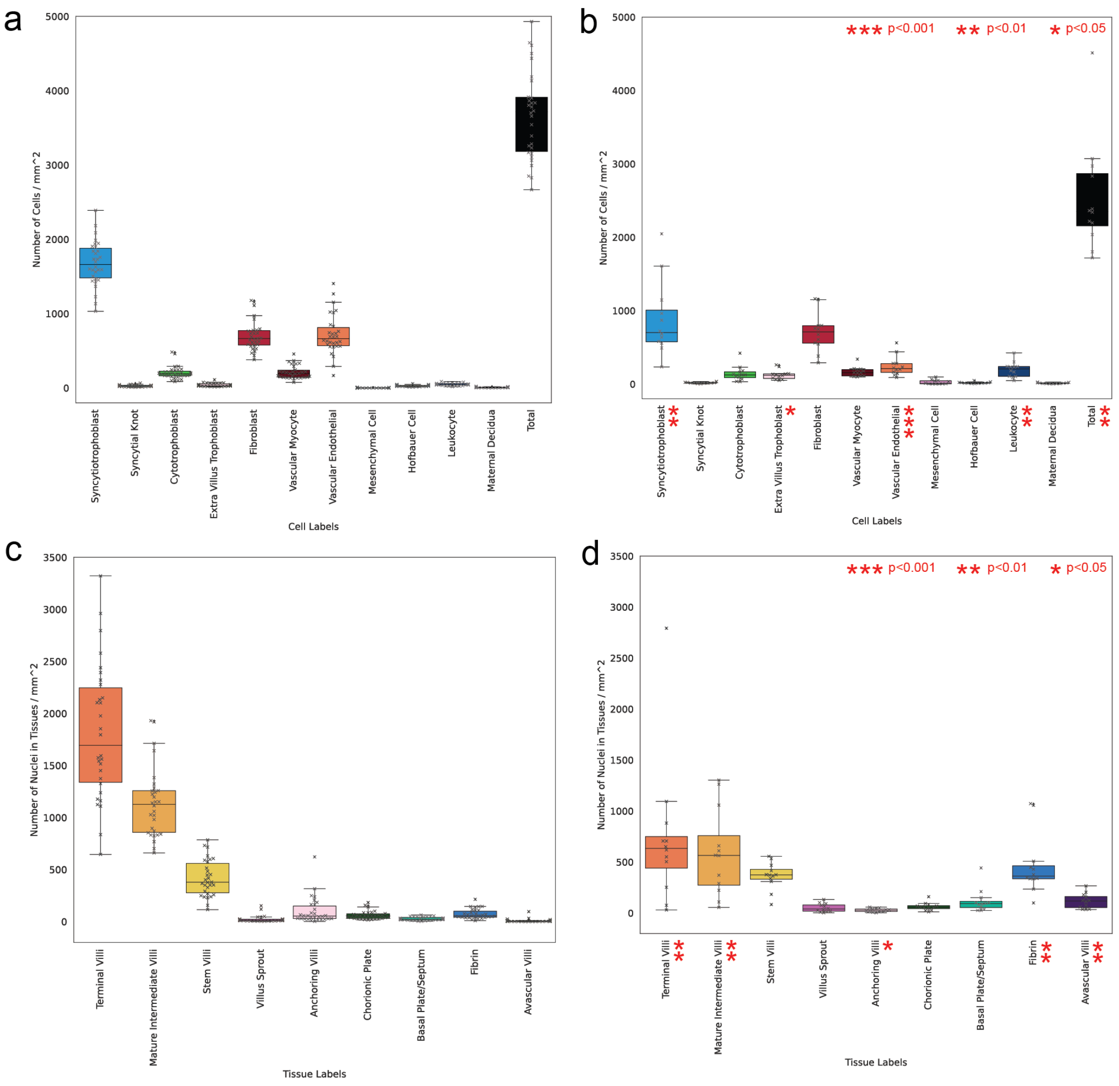


Figure 6.4: Predicted cell and tissue microstructure density per area mm^2 across 30 whole slide images (WSI) of healthy term placentas and 12 WSIs of placentas with placental infarction. (a) Cell density across healthy term placentas including total nuclei, (b) cell density across term placentas with placental infarction including total nuclei, (c) tissue node density across healthy term placentas, (d) tissue node density across term placentas with placental infarction. The box centre line represents the median, and whiskers are drawn up to 1.5 times the interquartile range. Each WSI datapoint is shown by a cross marker. Bonferroni adjusted significant differences in cell and tissue structures between the two groups, which were calculated using two-sided Welch's t-test are shown by red asterisks.

6.4.2 Lesion Whole Slide Image Data

The full parenchymal lesion dataset used in this chapter contains slides from 103 singleton placentas from 2nd trimester to term from two institutes. Placentas were selected as having one of the four parenchymal lesions of interest in pathology reports, and no placentas were excluded. As pathology reports are on a placenta level rather than slide level and not all slides will contain the lesion, I manually assessed each slide to ensure it contained the reported lesions. A secondary assessment was then given by a perinatal pathologist at another institute to confirm the lesion was present on the slide, and this assessment was used as the ground truth. In both cases, apart from knowing that placentas were selected to contain one of the four parenchymal lesions, assessors were blind to the lesions as stated in pathology reports. It was also noted whether the slide had small villi for gestational age, villous edema, or an inflammatory response. In total, 68 slides from UoT and 248 slides from HMC were included. See Figure 6.5 for the range of gestational ages from each institute and the number of slides with each lesion either as the only lesion on the slide or in combination with other lesions.

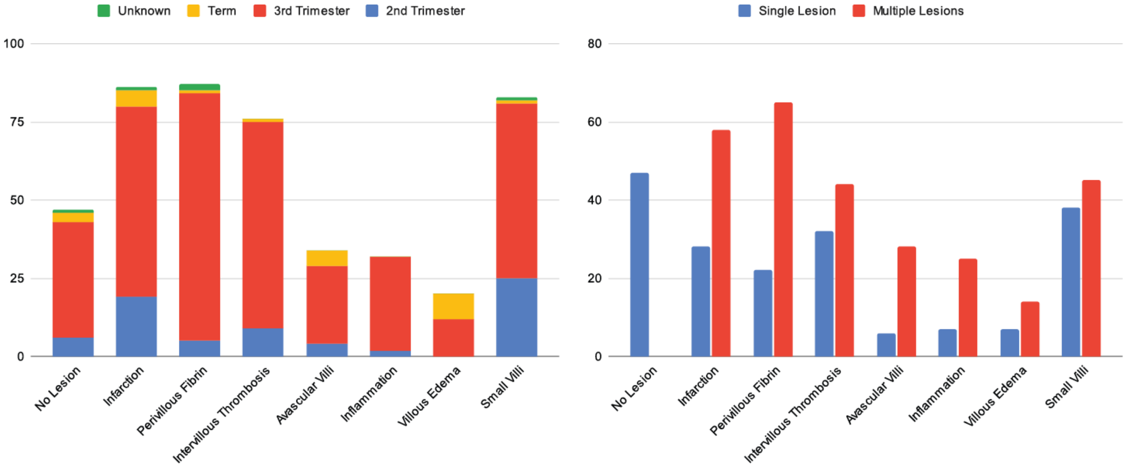


Figure 6.5: Number of slides with each parenchymal lesion by perinatal pathologist secondary assessment. The leftmost plot shows how lesions vary across gestational age in the data, with the majority of lesions from 3rd trimester placentas. The rightmost plot shows how many slides contain the lesion alone or are present with other lesions on the slide.

6.4.3 A Parallel Cell and Tissue Microstructure Prediction Workflow

The cellular and tissue microstructure predictions across whole slides are saved to a SQLite database. This has the benefit of providing powerful SQL queries across patient, slide, and completed inference run tables. As it isn't a client/server database engine, it has fast input/output for reading and writing operations, which is essential for scaling deep learning to whole slide inference, where each slide can contain over a million nuclei. However, a downside of not having a client/server database engine is the lack of concurrency support, especially in a computing environment with a distributed network file system. Slide inference needs to be parallelised to scale across hundreds of whole slide images.

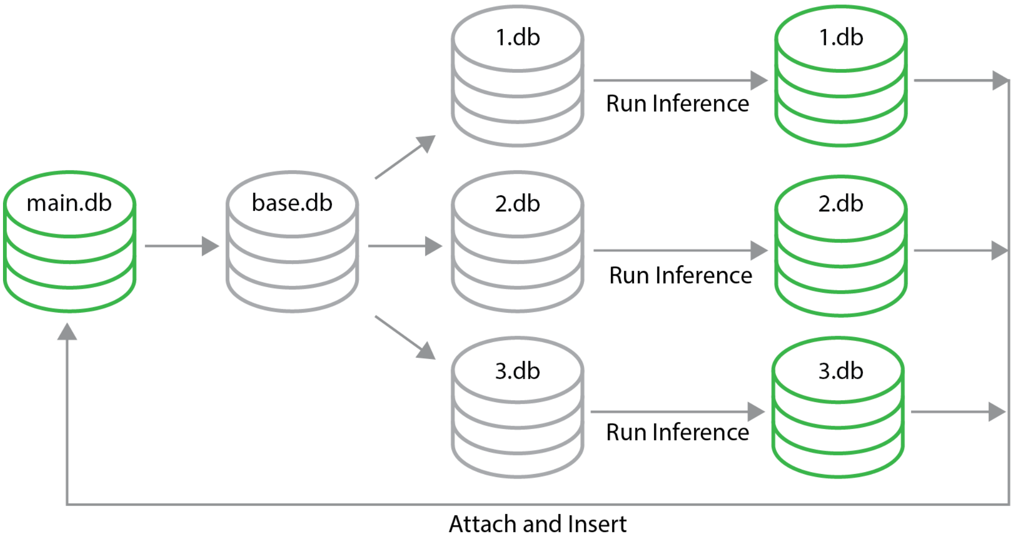
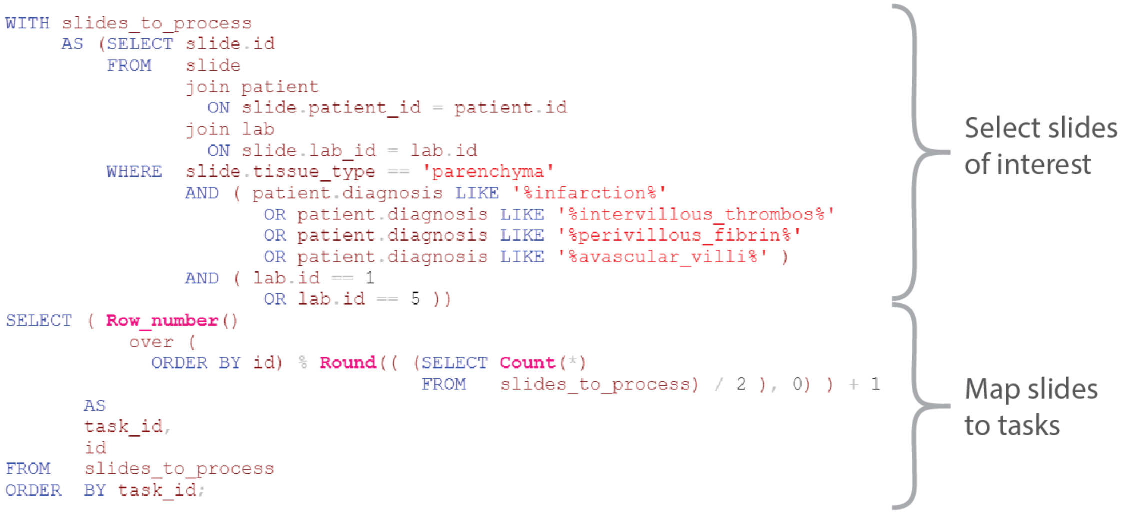


Figure 6.6: SQLite database workflow for running whole slide inference in parallel. From the source of truth dataset *main.db* a database *base.db* with metadata for running inferences is created. This is used to seed a task-specific database *{task_id}.db* for which model inference can be run in parallel across servers. After whole slide inference is complete, the resulting tables are stitched back into *main.db*.

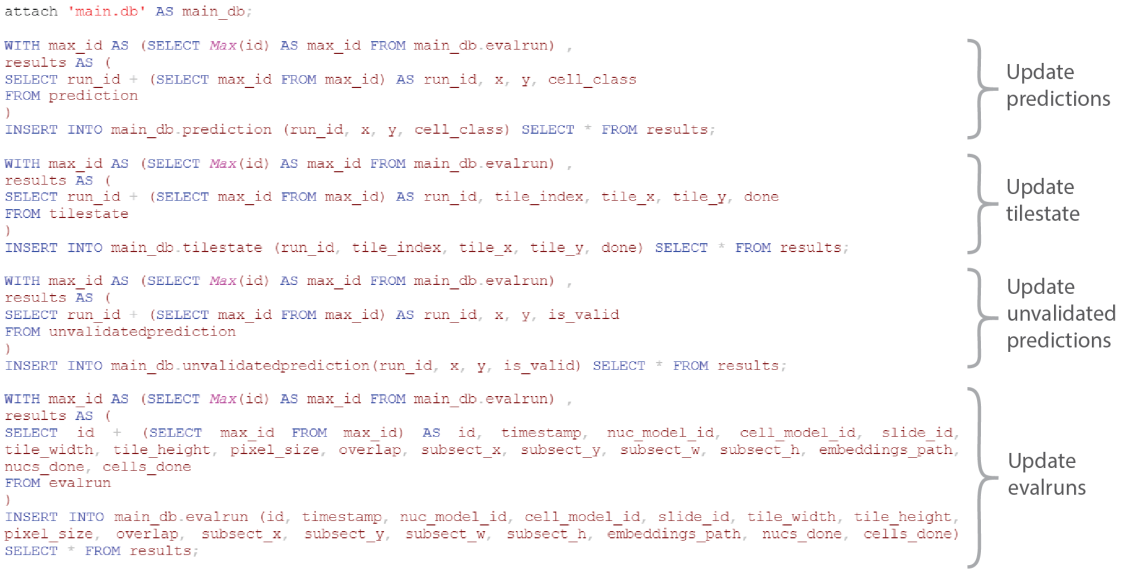
I developed a workflow to parallelise inference of the cell and tissue pipeline across multiple slides while ultimately still recording to the same SQLite database (Figure 6.6). The workflow centres around having a main source of truth database *main.db*, a metadata seeding database *base.db* and multiple temporary task-specific databases *{task_id}.db* which can be written to in parallel. A copy of *main.db* called *base.db* is first created containing all slide and model tables necessary for inference runs. As *main.db* can become very large over time, this saves disk space and avoids unnecessary copying of information. Depending on the specifics of the computing environment, the next step is to export the result of an SQL query to select relevant slides and partition them into tasks. If the computing environment allows for 2 full inference runs per task, then each task is mapped to two slides. The SQL query to select the placenta slides with lesions analysed in this chapter is shown in Figure 6.7.



```
WITH slides_to_process
     AS (SELECT slide.id
         FROM   slide
                join patient
                  ON slide.patient_id = patient.id
                join lab
                  ON slide.lab_id = lab.id
         WHERE  slide.tissue_type == 'parenchyma'
                AND ( patient.diagnosis LIKE '%infarction%'
                       OR patient.diagnosis LIKE '%intervillous_thrombos%'
                       OR patient.diagnosis LIKE '%perivillous_fibrin%'
                       OR patient.diagnosis LIKE '%avascular_villi%' )
                AND ( lab.id == 1
                       OR lab.id == 5 ))
SELECT ( Row_number()
           over (
             ORDER BY id) % Round(( (SELECT Count(*)
                                     FROM   slides_to_process) / 2 ), 0) ) + 1
       AS
       task_id,
       id
FROM   slides_to_process
ORDER  BY task_id;
```

Figure 6.7: SQL query for selecting placenta lesion slides and mapping them to tasks. After selecting the slides of interest, in this case, placenta parenchyma slides with one of the four parenchymal lesions, the query will return a table which maps two of these slide IDs to a unique task ID.

The task-slide mapping is then used to parallelise each inference run. For each task, a task database *{task_id}.db* is created from *base.db* and all results across slides within the task are recorded to that database. As there are no other processes writing to that task database, there are no concurrency concerns. After all tasks are complete, a bash script will check for any slides within a task that did not finish full inference across the slide. Any runs that were not completed may be restarted and will continue inference from where they stopped. Finally, the results in each task database are copied into *main.db* by attaching the two databases together and inserting them into each table at the next available run ID (Figure 6.8). Each task database can then be deleted. This workflow, predominantly formed of a series of bash and SQL scripts, has been used for the placenta lesion slides analysed in this chapter but also for adipose and liver histology by others in the lab.



```sql
attach 'main.db' AS main_db;

WITH max_id AS (SELECT Max(id) AS max_id FROM main_db.evalrun) ,
results AS (
SELECT run_id + (SELECT max_id FROM max_id) AS run_id, x, y, cell_class
FROM prediction
)
INSERT INTO main_db.prediction (run_id, x, y, cell_class) SELECT * FROM results;

WITH max_id AS (SELECT Max(id) AS max_id FROM main_db.evalrun) ,
results AS (
SELECT run_id + (SELECT max_id FROM max_id) AS run_id, tile_index, tile_x, tile_y, done
FROM tilestate
)
INSERT INTO main_db.tilestate (run_id, tile_index, tile_x, tile_y, done) SELECT * FROM results;

WITH max_id AS (SELECT Max(id) AS max_id FROM main_db.evalrun) ,
results AS (
SELECT run_id + (SELECT max_id FROM max_id) AS run_id, x, y, is_valid
FROM unvalidatedprediction
)
INSERT INTO main_db.unvalidatedprediction(run_id, x, y, is_valid) SELECT * FROM results;

WITH max_id AS (SELECT Max(id) AS max_id FROM main_db.evalrun) ,
results AS (
SELECT  id  +  (SELECT  max_id  FROM  max_id)  AS  id,  timestamp,  nuc_model_id,  cell_model_id,  slide_id,
tile_width, tile_height, pixel_size, overlap, subset_x, subset_y, subset_w, subset_h, embeddings_path,
nucs_done, cells_done
FROM evalrun
)
INSERT INTO main_db.evalrun (id, timestamp, nuc_model_id, cell_model_id, slide_id, tile_width, tile_height,
pixel_size,  overlap,  subset_x,  subset_y,  subset_w,  subset_h,  embeddings_path,  nucs_done,  cells_done)
SELECT * FROM results;
```

Figure 6.8: SQL query for updating the main database from a task database. For each task database, this query will attach the main database, find the last inference run (*evalrun*) id and update the *predictions*, *unvalidatedpredictions*, *titestate*, and *evalruns* tables. After this query has been run for each task database, the main database will look the same as if everything had been written directly to it.

6.4.4 Cell and Tissue Microstructure Differences

I compare the predicted cell and tissue microstructure densities for slides in each lesion group against the 30 healthy term placentas of the previous chapter. In the interest of space, I report only the classes for which two-sided Welch's t-tests are significant at values $p<0.001$ after Bonferroni multiple testing correction. We see from Tables 6.1 and 6.2 that for all four lesions, there are significant differences for vascular endothelial cells. Leukocytes and total density are also significant for all lesions apart from avascular villi. For avascular villi, total density is not expected to be very different as avascular villi are clinically significant in small quantities, and trophoblasts will still be present. In terms of tissue microstructures, all four lesions have significantly different densities for fibrin and predicted avascular villi in addition to mature intermediate villi. These results highlight the utility of cellular- and tissue-level analysis for slide-level understanding.

Table 6.1: P-values for cell density differences for each lesion of interest against the healthy group. P-values are calculated using a two-sided Welch's t-test with Bonferroni multiple testing correction, and only values where $p<0.001$ are reported.

	Infarction	Perivillous Fibrin	Avascular Villi	Intervillous Thrombosis
Syncytiotrophoblast	<0.001	-	-	<0.001
Extravillus Trophoblast	<0.001	<0.001	<0.001	-
Fibroblast	<0.001	<0.001	-	<0.001
Vascular Endothelial Cell	<0.001	<0.001	<0.001	<0.001
Mesenchymal Cell	<0.001	<0.001	-	<0.001
Leukocyte	<0.001	<0.001	-	<0.001
Total	<0.001	<0.001	-	<0.001

Table 6.2: P-values for tissue microstructure density differences for each lesion of interest against the healthy group. P-values are calculated using a two-sided Welch's t-test with Bonferroni multiple testing correction, and only values where $p<0.001$ are reported.

	Infarction	Perivillous Fibrin	Avascular Villi	Intervillous Thrombosis
Mature Intermediate Villi	<0.001	<0.001	<0.001	<0.001
Terminal Villi	<0.001	<0.001	-	-
Villus Sprouts	<0.001	<0.001	-	<0.001
Fibrin	<0.001	<0.001	<0.001	<0.001
Avascular Villi	<0.001	<0.001	<0.001	<0.001

6.4.5 Lesion Localisation Using Unsupervised Clustering

I explore the unsupervised patch-based clustering method from Chapter 4 to see if lesions can be easily clustered from cell and tissue microstructure features. I combine the cell and tissue microstructure predictions for each nucleus in a patch into proportions and use k-means clustering with 5 clusters to group patches. I randomly select three slides for each of the four lesions of interest, where that lesion is the only lesion present on the slide, as reported by secondary assessment. As before, patches are 200x200 µm, patches with fewer than 10 nuclei are excluded, and patch classes are assessed by manual inspection of cellular and tissue microstructure compositions.

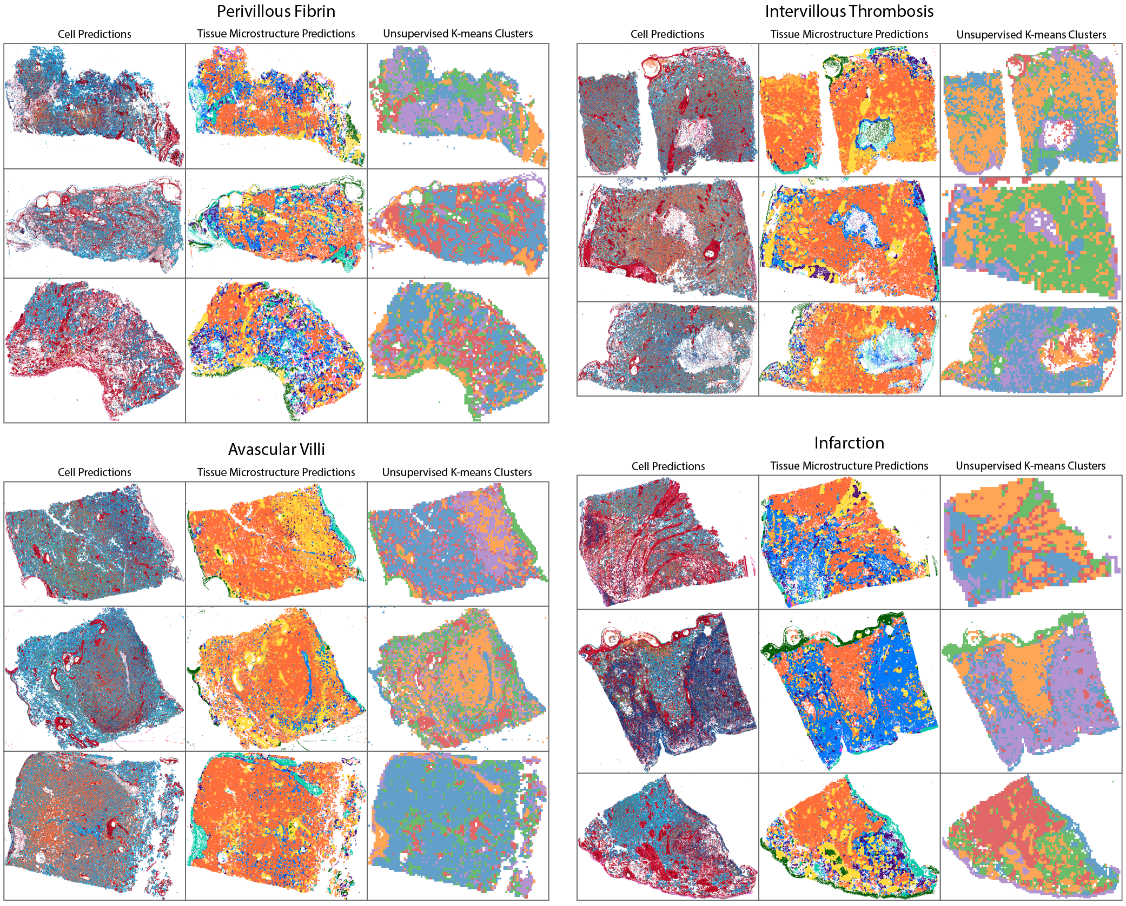


Figure 6.9: Cell, tissue microstructure, and k-means clusters on patches for slides with parenchymal lesions. For each lesion, slides are selected to be the only lesion on the slide. Patches are able to separate regions containing the lesion when there are large changes to the cellular and tissue compositions of those regions.

Figure 6.9 shows that when a lesion causes large cellular and tissue changes to the slide, the clustering approach distinguishes the lesion more clearly. This is the case for infarction and intervillous thrombosis and does not depend on the severity or size of the lesion. For avascular villi, when the lesion is severe, as is the case with the two topmost slides of the figure, the clustering method identifies the lesion even when this isn't directly obvious from the macro-scale cell and tissue microstructure predictions. However, a lesion patch class is not readily identifiable when avascular villi are present in smaller quantities, such as the bottom-most slide. For perivillous fibrin, lesion patch classes on fibrinoid regions seem to follow the tissue microstructure prediction of fibrin. As before, this analysis would benefit from a quantitative comparison to regions with lesions annotated by a perinatal pathologist. However, it suggests that the resulting changes to the cellular and tissue microstructure compositions can be used to identify parenchymal lesions.

6.5 Section 2: Automated Lesion Classification and Localisation

In the previous section, I showed how, in principle, lesions can be identified on a slide by their resulting cellular and tissue changes. However, while an aggregated slide-level analysis may indicate that there is a lesion on the slide, it does not tell us where that lesion is on the slide. The unsupervised k-means clustering approach can identify unusual regions which may correspond to the lesion, but the model is only applied on a slide-by-slide basis with no ability to learn from lesions across samples. A supervised approach could classify, segment the lesion, and train across all slides but would require manual lesion-specific annotations. In the following analysis, I explore a self-supervised method trained across all slides that uses the underlying cell and tissue information to cluster regions, thereby classifying and localising lesions. Conceptually, it is an extension of the k-means approach applied across slides with clustered regions not fixed to patch boundaries.

I explore a novel graph compression autoencoder (GCAE) across the lesion slide dataset using cellular and tissue microstructure embeddings. At a high level, the goal is to use an encoder-decoder architecture to compress the whole slide graph into an iteratively smaller graph representation, the encoded features of which summarise regions across the original graph. The decoder aims to reconstruct the original cellular and tissue microstructure node features, and the whole model is trained end-to-end based on this reconstruction performance. This training scheme encourages the encoder to meaningfully summarise the graph at each compression, which acts as a bottleneck so that the decoder can reconstruct from these compressed representations. The encoded features can then be used for downstream analysis of regions exhibiting cellular and tissue microstructure changes, potentially due to a lesion.

6.5.1 Related Work

In the deep graph learning space, unsupervised graph autoencoders and variational autoencoders [287] have been explored for edge reconstruction [287, 288] and node feature reconstruction [289–291]. Different types of encoding strategies include edge masking [287], node masking [289, 290] and feature corruption [291]. Decoders range in complexity from inner dot product [287], which compares latent variable similarity for edge reconstruction to fully trainable GNNs [288–291]. Few methods combine graph compression and autoencoders [292, 293] but are supervised and do not scale to large graphs [294].

Graph compression, or graph pooling, is most commonly used for graph classification. Here, a graph and its node features are iteratively compressed into smaller graph representations until a single vector remains, which predicts the graph class. The selection of supernodes for pooling can be learnt [292, 295–299] or calculated algorithmically [300–302]. Graph classification is unlikely to work on the lesion dataset due to the small dataset size (<400 graphs) and multi-label task, likely leading to overfitting and imbalanced class prediction.

Whilst not an autoencoder structure, the only application of graph compression for graphs in histology are from [214] using DiffPool [298] and [219] using MinCut pool [299] for graph classification and node clustering, respectively. As these methods can hierarchically pool the graph, their feature representations can be queried at different graph levels, effectively summarising the graph. However, both DiffPool and MinCut pool only support a dense adjacency matrix rather than the more common and memory-efficient sparse adjacency matrix and so cannot be used with large graphs (not to be confused with sparse or dense graphs).

The methodology presented here applies iterative graph compression and reconstruction to whole slide graphs. This approach takes inspiration from supervised methods such as Graph U-Net [292], without skip connections, and from the hierarchical compression and interpolation of methods for point cloud analysis such as PointNet++ [300] and PointTransformer [303].

6.5.2 Methodology

Dataset Split

For training and evaluating the GCAE, I split the lesion-stratified slides into training, validation, and test sets (Figure 6.10). The test set, forming 11% of the data, was first manually and deliberately constructed to contain a representative set of slides with both single and multiple lesions. This was made to ensure that each lesion had representative slides at differing severities and in combination with each other. The training and validation sets were then randomly split across the remaining slides into 74% and 15% sets.

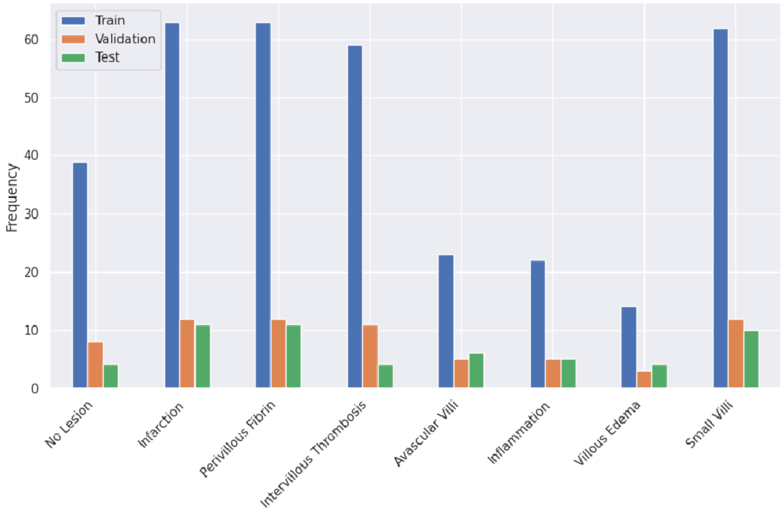


Figure 6.10: Dataset split across the lesion stratified slides. Lesions of interest are predominantly infarction, perivillous fibrin, intervillous thrombosis and avascular villi, but secondary assessment annotations for inflammation, villous edema, small villi and no lesion are also included.

Graph Construction

The whole slide graph for each slide is constructed as in Chapter 4 with a few key differences. Min-max normalised edge distances are included as 1-dimensional features on edges, and node features are the concatenated 64-dimension cellular and 64-dimension tissue microstructure embedding vectors for a total feature vector of 128 (Figure 6.11). The tissue microstructure prediction GNN is slightly tweaked to match the cell classification model's linear layers. The previous single 64-dimension embedding extraction linear layer is replaced with two linear layers of 128 dimensions and 64 dimensions with ReLU activation only after the first layer and batch norm [304] after both. This mirrors the custom linear layer structure of the cell classification model and avoids only positive and zeroed-out values from ReLU. As before, graph normalisation is applied independently to the cell embeddings, but here, it is additionally applied to the tissue microstructure embeddings such that both sets of embeddings have a mean of 0 and a standard deviation of 1. Finally, at each training epoch, the whole slide graph is randomly sampled to half the size (effectively a node dropout of 50%) to reduce GPU memory requirements and add regularisation. Edges are reconstructed using a k-NN graph with k=6.

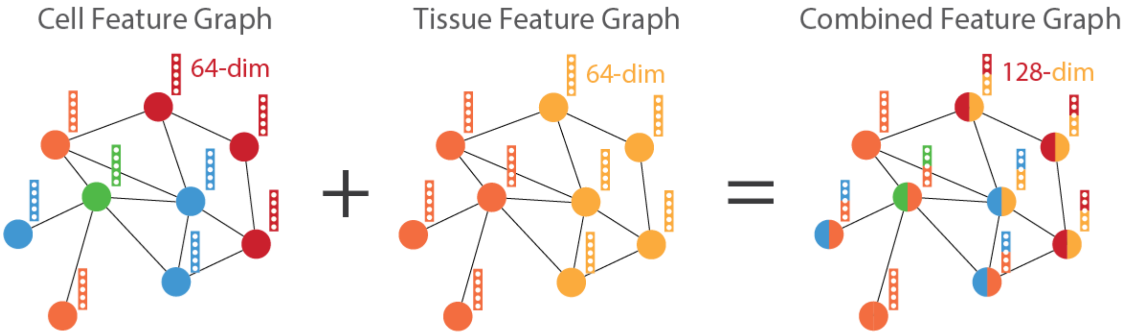


Figure 6.11: Cell and tissue microstructure graphs are combined into one whole slide graph. Both cell and tissue graphs share the same nuclei nodes and node coordinates, and each contains 64 dimension node features extracted from the previous prediction models. Both feature sets are standardised to have a mean of 0 and a standard deviation of 1 before being concatenated into new 128-dimension node feature vectors.

Compression Graph Autoencoder Model Architecture

As with all autoencoders, the model consists of an encoder which iteratively removes information from the input data and a decoder which aims to reconstruct the original information, or some other desired output, from the encoded data [134, 287, 305]. In my case, I use an encoder to iteratively compress the graph topology into a smaller graph representation and a decoder to reconstruct the features of the original graph given the graph topology and compressed node features. I use a childnode/supernode structure to model the relationship between nodes across compression depths. The nodes in the input graph are the topmost childnodes, and nodes in the final compressed graph are supernodes of all other nodes. At a high level, the goal of the encoder is to learn how to aggregate cell and tissue regions into meaningful vectors representing said region, leading to each compressed graph summarising the graph at the previous depth. The goal of the decoder is to learn how to use these summarised vectors and their surrounding topological information to reconstruct the surrounding node features. The encoded embeddings from the most compressed graph can then be clustered and projected back through the matching childnodes to the input graph to localise different regions, including regions containing lesions.

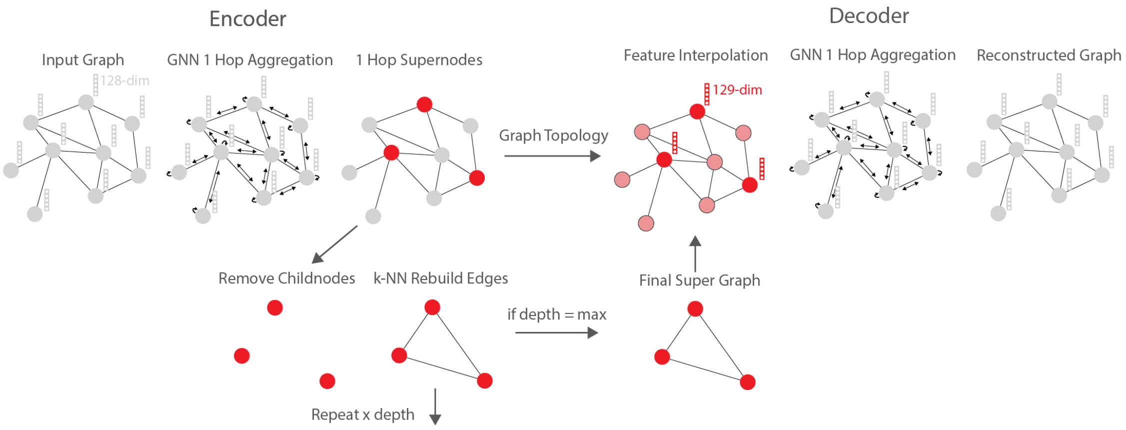


Figure 6.12: The autoencoder architecture for graph compression and feature reconstruction. At each depth, the encoder takes a graph, applies one hop GCN aggregation, selects supernodes and removes their one hop neighbours before reconstructing new edges for the compressed graph. Once the maximum depth is reached, the compressed graph is passed to the decoder. At each depth, the decoder interpolates the input graph's node features to nodes in the next graph topology before applying a one hop GNN aggregation. The model is trained end-to-end for node feature reconstruction.

Encoder. The encoder consists of node feature aggregation and one hop neighbourhood compression (Figure 6.12). A single GNN layer at each depth first aggregates the one hop neighbourhood of each node, effectively compressing that node's feature vector with its neighbourhood. After aggregation, a custom supernode selection algorithm iteratively selects supernodes and removes their one hop neighbours. From the supernodes, a new graph is constructed using k-NN edges with k decreasing by 1 at every depth from a starting k=6. This is repeated until the maximum depth is reached. A single GNN layer is updated end-to-end from the final reconstruction loss at each depth.

The one hop compression algorithm iteratively selects supernodes such that between 65-75% of nodes are removed from the graph at each depth, all childnodes are one hop neighbours of supernodes, and no supernodes are one hop neighbours of each other (Figure 6.13). The algorithm first defines an iteration size (1000) and reduction factor (0.75) that exponentially decays by an increasing factor of 1 at each depth. For all unassigned nodes, nodes are randomly selected as candidate supernodes up to the iteration size. If there are fewer unassigned nodes than the iteration size times by the reduction factor, then the iteration size is reduced by the reduction factor and supernode selection is repeated. In each iteration, any selected supernodes which are one hop neighbours of each other have one of those supernodes removed from the candidate supernodes. All one hop neighbours of the remaining selected supernodes are assigned to childnodes. This process is repeated until all nodes are either supernodes or childnodes or the iteration size is 1. All steps are performed as vector operations on device for efficiency.

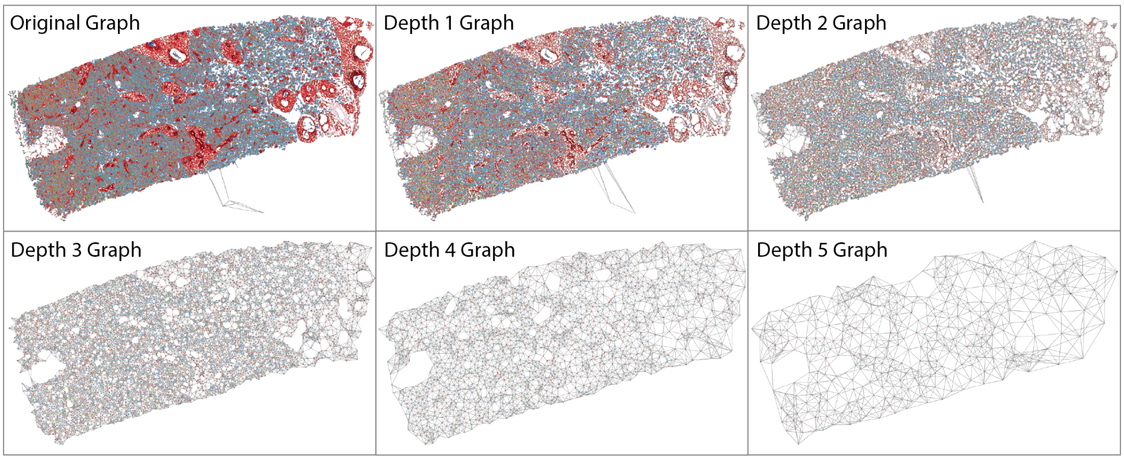


Figure 6.13: Resulting graphs from one hop compression across depths coloured by original cell predictions on nodes. An input graph with around 1 million nodes will have around 1000 nodes by the 5th compression depth.

This one hop compression algorithm ensures that the one hop GNN aggregation includes all childnodes of every supernode in the aggregated neighbourhood. It will result in a different supernode selection and compressed graph at every epoch, providing regularisation during training. It is a 10 times faster supernode selection alternative to farthest point sampling (introduced by PointNet++ [300] and used by CGC-Net [214] (Table 6.3) whilst still achieving a more representative supernode spread than random supernode sampling (Figure 6.14).

Table 6.3: Comparison of total sampling time across 5 depths for two WSIs with over a million nodes each. Custom one hop compression is marginally slower than random sampling but is 10 times faster than farthest point sampling. The best result is shown in bold, and the technique used in this chapter is underlined.

Supernode Sampling Method	Time Across Two WSIs (s)
<u>One Hop Compression</u>	<u>18.34</u>
Farthest Point Sampling	182.53
Random Sampling	**15.02**

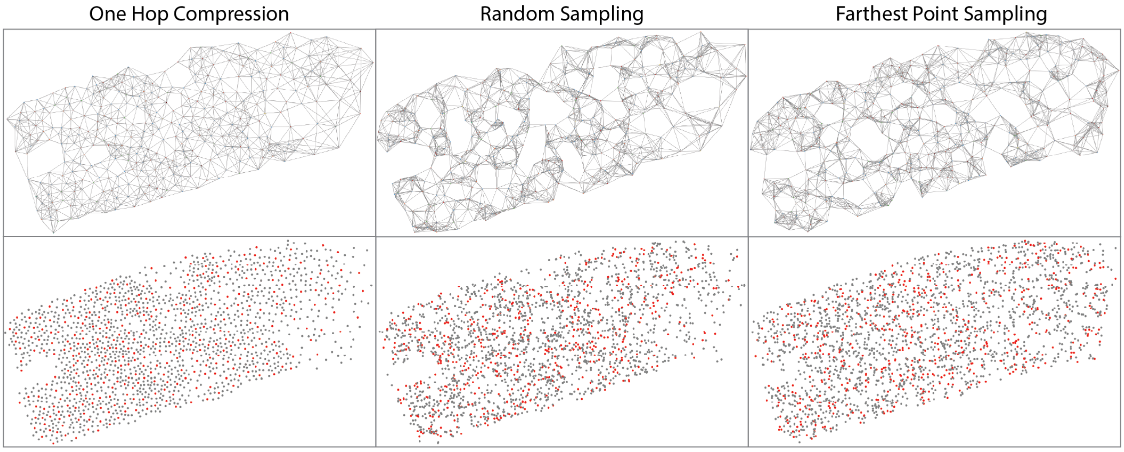


Figure 6.14: Graphs at the lowest depth of 5 and supernode to childnode relationship between depths 4 and 5 for one hop compression, random sampling, and farthest point sampling. One hop compression results in a more even distribution of selected supernodes across the graph compared to random sampling and farthest point sampling. Supernodes (shown in red) are not direct neighbours of each other and are surrounded by their one hop childnodes (shown in grey).

Decoder. The decoder consists of feature interpolation and node feature aggregation for reconstruction (Figure 6.12). At each depth, the decoder is given a compressed graph with node features and the graph topology of the next uncompressed layer (Figure 6.15). An additional node feature of the node's node degree is added as the 129th feature. The inclusion of node degree and edge distance weights should provide the model with surrounding topological information when decoding features. For example, to reconstruct terminal villi embeddings, the model may consider that terminal villi nuclei are found in closer proximity to other nuclei and in denser regions compared to other tissue microstructures.

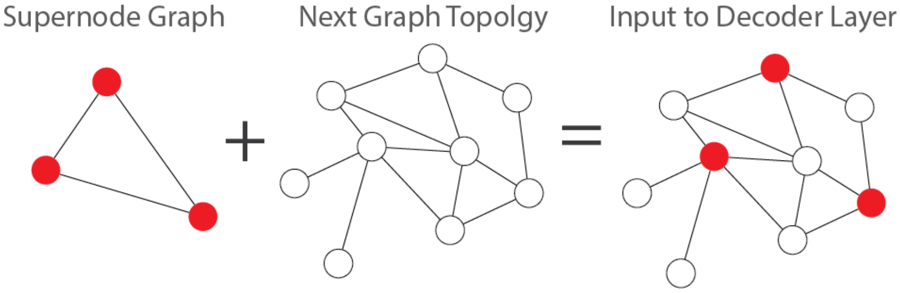


Figure 6.15: The two inputs to each decoder layer. At each layer, the decoder is given a supernode compressed graph with supernode features and the graph topology of the next level for which the decoder must reconstruct features.

The decoder first performs k-NN interpolation [300], which uses the inverse of weighted distances, based on k nearest neighbours (k=6), to assign the weighted mean of features from supernodes to nodes in the next graph topology. This has the effect of spreading the supernode features relative to their distance across the new uncompressed childnodes. After interpolation, a single-layer GCN uses the one hop neighbourhood of each node to aggregate and assign new node features. This is repeated until the original input graph is reached, at which point node similarity is compared against the original nodes for a reconstruction loss. As with the encoder, a single GNN layer learns the aggregation at each depth and is updated end-to-end from the final reconstruction loss.

Encoded Embeddings. The embeddings from the most compressed (supernode) graph are clustered using k-means clustering into 4 clusters. These cluster labels are then projected iteratively through the compressed graphs to the original input graph. In the interest of time and memory, the exact childnode to supernode relationship is not stored and is recalculated at this step by finding the one hop neighbourhood from supernode to childnodes at the next level. A single supernode to childnode pair is formed at each level until the smallest and largest graphs are paired. See Figure

6.16 for random class assignment of supernodes and their projection to childnodes in the original graph. In practice, childnodes belong to multiple supernodes (and clusters), which should be accounted for in future work.

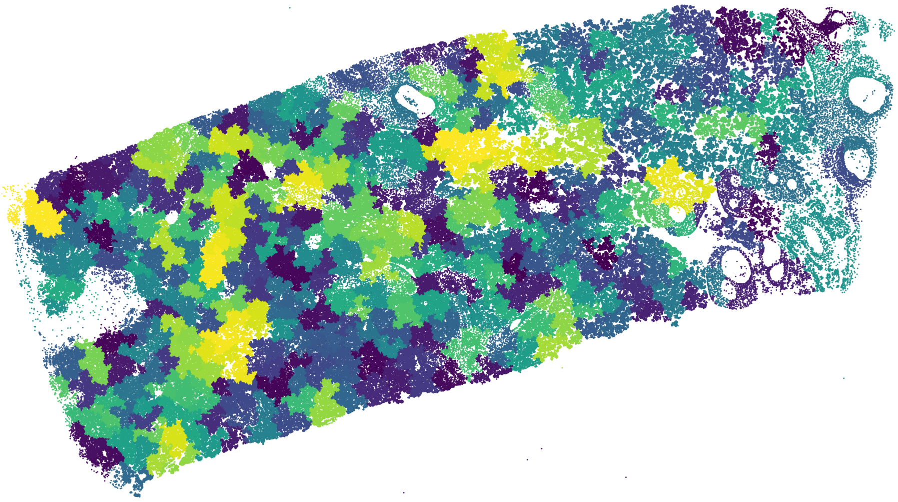


Figure 6.16: Random class assignment to supernodes and their projection across compression depths to childnodes. This projection roughly demonstrates the supernode to childnode relationship from the most compressed to the original graph. In practice, childnodes belong to more than one supernode, which should be accounted for in the downstream task.

Training Scheme

The GCAE is trained for 165 epochs with a batch size of one whole slide graph, 0.001 learning rate, an Adam optimiser [188], and a Mean Squared Error (MSE) loss on the reconstruction of the original 128 node features. The model has a compression depth of 5 after the initial 50% random sampling, the GNN layer at each depth has 128 hidden units in the encoder and 129 hidden units in the decoder to account for node degree, and all GNN layers are followed by ReLU activations. The final reconstruction is performed by a single linear layer with 128 hidden units.

Three GNN layer types are explored: Graph Convolutional Network (GCN) [118], Graph Attention Network V2 (GATv2) [261] with one attention head, and Graph Isomorphism Network (GIN) [306]. As one of the first message-passing architectures, GCN offers a simple aggregation scheme to use as a benchmark. GATv2 introduces learnt attention to weight the neighbourhood aggregation, and GIN is a more expressive model that uses an internal MLP to compute its aggregation. All layer types are used with their variants that support edge weights. A randomly initialised GATv2 is used as a baseline, and further comparisons are made to a GIN model without k-NN interpolation [300] and a GIN model with tanh activations in the decoder and at final encoder depth so that latent features more closely match the target input features. Depending on the graph size, each model takes 2-14 seconds per whole slide graph on an A100 NVIDIA 80GB GPU.

6.5.3 Model Evaluation

As we do not have ground truth annotated regions containing lesions against which to compare predictions, I assess model performance in three ways. I first quantitatively compare reconstruction performance between models to select the best-performing model type. Using this model, I further assess reconstruction performance by qualitatively comparing cell and tissue predictions from these embeddings against the original cell and tissue predictions. Finally, I explore the quality of the encoded supernode features for downstream region representation.

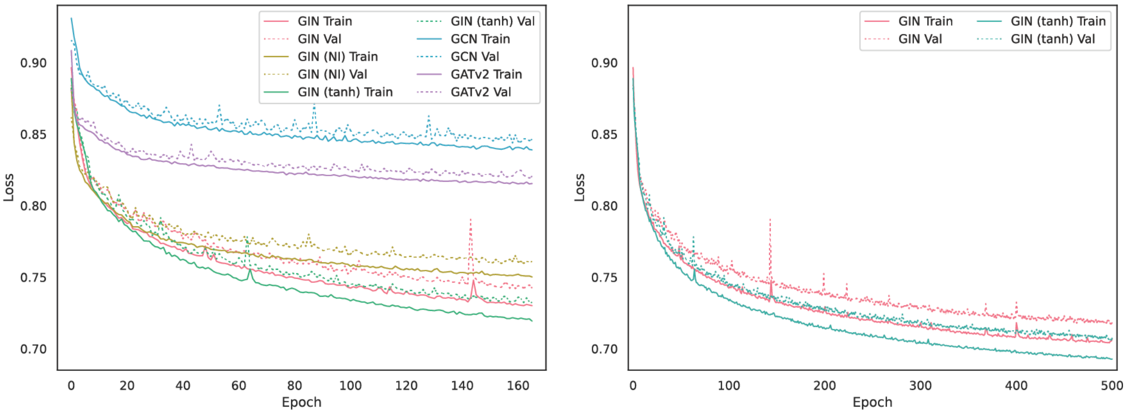


Figure 6.17: MSE loss curves during training on train and validation data for all explored models. Variants of GIN, the most expressive model of the three types, have the lowest MSE loss across epochs. GIN (NI) denotes the GIN model trained without k-NN interpolation in the decoder and is marginally worse than the GIN model with k-NN interpolation. GIN (tanh) denotes the GIN model trained with tanh decoder activations rather than Relu and is the best-performing model.

In Figure 6.17, I show the mean node MSE loss at each epoch during training on train and validation data. We see that GIN outperforms all other GNN layer types and has not yet overfit to the training data, showing the potential for further training. Of the GIN variants, using k-NN interpolation and GIN with tanh activations in the decoder shows the best training and validation performance across epochs. We see a similar pattern when we continue to train the two best GIN models to 500 epochs. GIN with tanh activations outperforms GIN with Relu activations, and neither model has fully converged by 500 epochs.

In Tables 6.4 and 6.5, I take the model with the lowest validation MSE for each model type and evaluate its cell and tissue reconstructions using Huber loss (δ=1.0), cosine similarity and Pearson Correlation Coefficient (PCC) on test data. Huber loss (or Smooth Mean Absolute Error) is similar to MSE but calculates outliers higher than a delta value using Mean Absolute Error, making it less sensitive to those outliers. Cosine similarity calculates the orientation between vectors in high dimensions, where -1 is the opposite direction and 1 is the same direction, making it insensitive to scale. PCC, also ranging from -1 to 1, measures the linear correlation between values and, thus, how well a linear function could describe the differences between the data. As with MSE loss during training, we see that GIN with k-NN interpolation and tanh activations outperforms all other GNN layer types up to 165 epochs with marginal improvements from training to 500 epochs. All GNN layers outperform the random baseline, but the results indicate that reconstruction performance can be further improved.

As the GCAE output features are the reconstructed embeddings of penultimate layers from prior models, we can pass the reconstructed features through the output layers of those models and visualise the resulting cellular and tissue microstructure predictions. I take the reconstructed features from the best performing model, GIN with k-NN interpolation and tanh activations at 500 epochs, and unstandardise them by applying the inverse of the standardisation function to both 64 dimension feature sets. I then contrast the reconstructed predictions against the original predictions in Figure 6.18. The reconstructed cell embeddings show an over-smoothing effect with a high predicted cell class homophily and underprediction of minority classes. Broadly, however, the predicted cells separate into the original macrostructure

Table 6.4: Performance of explored models for cell reconstruction across test data. Each model evaluated across test data is selected for its lowest MSE performance on validation data. GIN surpasses all other GNN layer types across all metrics for cell reconstruction. GIN without k-NN interpolation is denoted as GIN (NI), and GIN with tanh decoder activations as GIN tanh, with the number of epochs shown in brackets. The best results are highlighted in bold.

	Cell Huber ↓	**Cell Cosine ↑**	**Cell PCC ↑**
Baseline	0.438	-0.005	0.001
GCN	0.403	0.228	0.261
GATv2	0.404	0.219	0.253
GIN	0.374	**0.330**	0.372
GIN tanh	**0.373**	**0.330**	**0.374**
GIN (NI)	0.378	0.315	0.359
GIN (500 e)	0.369	0.341	0.385
GIN tanh (500 e)	**0.368**	**0.345**	**0.390**

Table 6.5: Performance of explored models for tissue reconstruction across test data. Each model evaluated across test data is selected for its lowest MSE performance on validation data. GIN surpasses all other GNN layer types across all metrics for tissue microstructure reconstruction. GIN without k-NN interpolation is denoted as GIN (NI), and GIN with tanh decoder activations as GIN tanh, with the number of epochs shown in brackets. The best results are highlighted in bold.

	Tissue Huber ↓	**Tissue Cosine ↑**	**Tissue PCC ↑**
Baseline	0.414	-0.032	0.000
GCN	0.324	0.402	0.472
GATv2	0.304	0.445	0.521
GIN	0.269	0.518	0.605
GIN tanh	**0.263**	**0.529**	**0.618**
GIN (NI)	0.281	0.493	0.580
GIN (500 e)	0.254	0.549	0.635
GIN tanh (500 e)	**0.247**	**0.562**	**0.650**

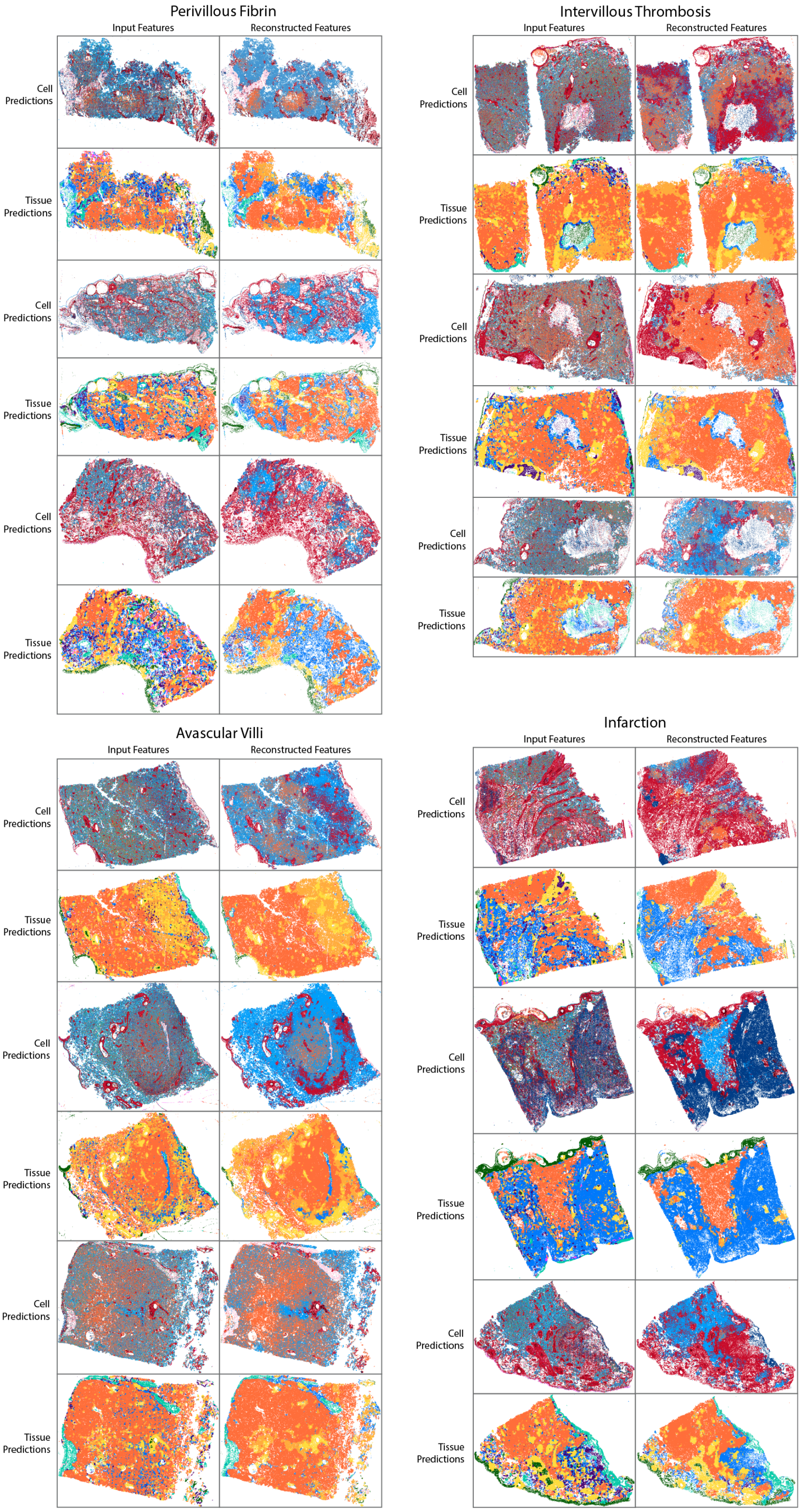


Figure 6.18: Original cell and tissue microstructure predictions compared to predictions from reconstructed features across slides with lesions. Predictions from reconstructed features match the original predictions on a macro scale.

regions. There is an overprediction of terminal villi for the tissue embeddings, but, as with the cells, the original regions can be identified. In terms of lesions, for infarction, intervillous thrombosis, and perivillous fibrin, the characteristic cellular and tissue microstructure changes are reconstructed, and the lesion can be visually identified from these changes on the slide. Regions of avascular villi are hard to distinguish on some of the slides from cellular and tissue microstructure changes alone, which is reflected in the reconstructed predictions. After unstandardising the input and reconstructed features, cell and tissue microstructure cosine similarity across these slides is 0.763 and 0.818, respectively.

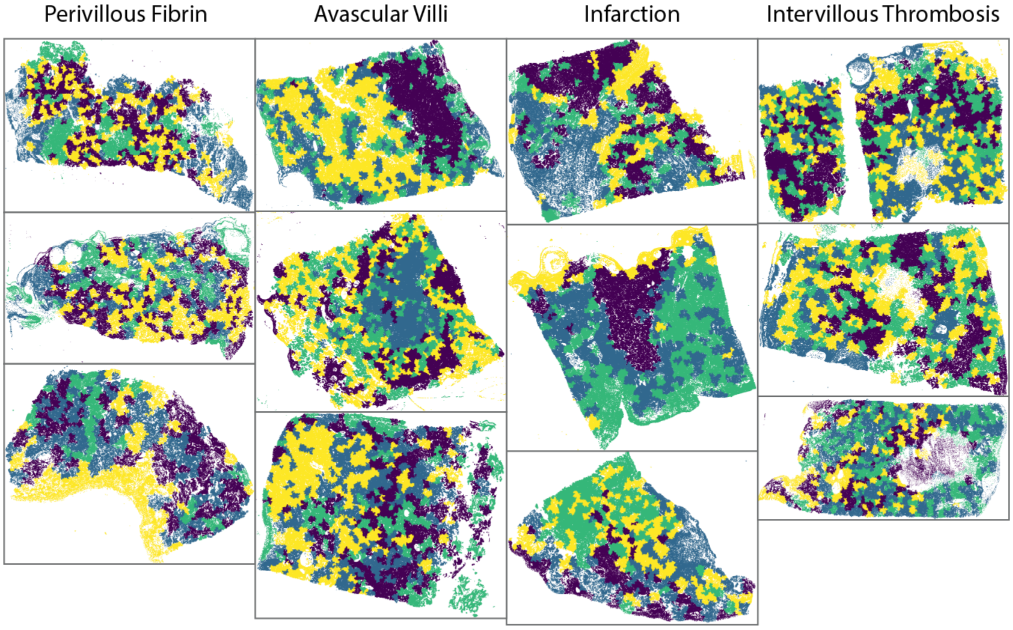


Figure 6.19: K-means clusters on supernode features projected to original nodes using a one hop traversal. After clustering the supernode features into 4 clusters, cluster labels are assigned to one hop childnodes by traversing through each uncompressed graph until the original graph is reached. Supernode cluster regions capture macrostructures and lesions to some extent, but there is further room for improvement.

Finally, I qualitatively evaluate the encoded supernode features for the downstream task of graph region summarisation. I use k-means with 4 clusters on the node features of supernodes and project those cluster labels to nodes in the original graph. With 4 clusters, I expect regions to be split into distal villi, stem villi and chorionic plate, basal plate (if present), and the regions containing a lesion. In Figure 6.19, I visualise the projected clusters on slides with a single instance of each of the four lesions. Comparing the original cell and tissue microstructure predictions in Figure 6.18 to these clusters, we see that projected supernode clusters summarise macrostructure regions to some extent, but there is some confusion among neighbouring areas. It is worth noting that the supernode to childnode projection technique used here does not account for shared childnodes and may not reflect the true supernode to childnode relationship. Nonetheless, the quality of supernode features may be improved with future work.

6.5.4 Discussion

This approach is a promising novel direction for unsupervised hierarchical graph clustering but needs further work. Firstly, this analysis would benefit from quantitative metrics to assess model performance on the downstream task. For example, perinatal pathologists could annotate lesions at a low resolution across representative or test dataset slides, and prediction overlap can be assessed using segmentation metrics such as Intersection over Union and Dice Coefficient. The models would also benefit from an ablation analysis, such as removing all GCN components to investigate a possible oversmoothing effect caused by k-NN interpolation, removing the node degree feature and not using edge weights. Using loss metrics that are less sensitive to outliers, like cosine similarity or a combination of losses, may improve performance. A general hyperparameter exploration of depth, learning rate, model size, GNN layer number, Delaunay Triangulation edges, etc., could result in higher-performing models. An embedding analysis using UMAP could also offer insight. Further things to consider are the lack of expressivity and learning flexibility of the one hop GNNs at each depth, that the current decoding structure could be leading to an oversmoothing of features, and how the class imbalance inherent to the cellular and tissue embeddings may be influencing the graph compression model to only reconstruct embeddings for the majority classes.

6.6 Conclusion

In this chapter, I have presented a preliminary exploration into lesion cellular and tissue phenotype quantification and unsupervised lesion classification and localisation. In the first section, I use the previous cell and tissue microstructure prediction models to assess aggregated whole slide changes across slides with parenchymal lesions. I find that there are significant differences in the cell and tissue microstructure type densities across slides and that these match the expected biological changes caused by lesions. This analysis shows that even using models trained for the healthy term placenta, whole slide aggregated cellular and tissue phenotypes can be expanded to lesion quantification. I additionally extended the previous unsupervised patch-based clustering method to include tissue microstructures and qualitatively found that lesion patches can be clustered by changes to cellular and tissue composition.

In the second section, I presented a novel graph compression autoencoder for unsupervised region clustering. The autoencoder includes a custom one hop compression algorithm that is 10 times faster than farthest point sampling, provides an even supernode selection across the graph, and ensures that a one hop GNN correctly aggregates each supernode's childnode neighbourhood. I explored three different GNN layer types for the encoder and decoder and found that GIN was the best performing layer type. However, further work is needed to improve supernode to childnode projection and supernode embeddings for downstream analysis.

In the context of the literature, this is the first automated method examining cellular and tissue microstructure phenotypes across whole slides for parenchymal lesion quantification. Further work, which includes training the cellular and tissue microstructure models on slides from earlier trimester placentas, is needed to use these metrics for lesion severity assessment. From a graph deep learning perspective, the presented graph compression autoencoder is a new architecture for feature reconstruction with a graph compression bottleneck, and this is the first application of graph autoencoders to histology cell graphs.

6.7 Limitations

This analysis was limited to four parenchymal lesions but could be extended to include all clinically relevant placental lesions. If the cellular and tissue microstructure models are updated to include classes from earlier gestational ages, then the autoencoder could be used to find regions showing accelerated or decelerated villus maturation. The cell model currently classifies leukocytes, but inflammatory lesions have not been explicitly explored. It may be challenging from existing model outputs to identify both the tissue microstructure of a region and that it is inflamed, so further model changes may be required.

Improvements may also be made to the lesion slide dataset. Currently, slides are only labelled as containing a lesion when it is above the clinically relevant threshold. However, this label could be improved to account for lesion severity, which would facilitate an exploration of how cellular and tissue microstructures vary by lesion severity. Although the lesion dataset included slides from 2nd and 3rd trimesters, the tissue microstructure model was only trained with tissues from term placentas. This means tissue microstructure predictions are likely biased towards the structures most common in a term placenta, such as terminal villi. Finally, additional healthy control slides are needed to improve the reliability of the whole slide cell and tissue microstructure comparison to slides with lesions.

Regarding the graph compression autoencoder, there was only a limited hyperparameter exploration in terms of GNN model type, and the approach would benefit from further methodological work and training across more slides. Current memory requirements of this training scheme require that the whole slide graph at each depth is loaded into GPU memory. This limits GPU hardware options and leads to a batch size of 1, which can cause large slide-specific changes to the model at each update step. A subgraph sampling approach for a graph autoencoder would be a novel contribution to graph learning. Finally, as with all unsupervised methods which use clustering for downstream analysis, an additional process will be needed to map unsupervised regions into lesion classes. This could be by fine-tuning with slide-level labels or through an embedding space similarity search [159, 164, 219].

Life must be understood backwards.
But it must be lived forward.

— Søren Kierkegaard

7

Future Work

Contents

7.1 Methodological Improvements

There are immediate, medium-term, and long-term ways to build upon the current work. As an immediate next step, the methods presented here can extend to the analysis of placentas from earlier trimesters, to include the three other placenta slide types, and to cover all lesions currently of clinical interest. Doing so would involve increasing the types of predicted cells (and cell-level features) and tissue microstructures. Models can be trained with data from more institutes to improve robustness, and whole slide metrics can be compared across larger slide datasets to establish reliable, healthy baselines and quantify lesion severity.

7.2 Automated Lesion Classification and Localisation

Following the last chapter, we can further improve the automated lesion classification and region localisation methodology. The ideal end goal is to have an unsupervised method which has explainable outputs by using identifiable biological features. The method should segment regions containing lesions on a slide and classify those lesions without manual inspection or heuristics.

There are a few variants to the existing methodology which could be explored first. The first variant is to backpropagate and update each model at each depth for that depth's reconstruction quality, either individually or in combination with the final reconstruction loss. Another variant would be introducing learning to supernode selection, perhaps by applying a score to candidate supernodes or learning the supernode-to-child relationship directly. However, this runs the risk of losing too much of the graph topology if, for example, all selected nodes are on one side of the graph and childnodes are not evenly distributed for aggregation. Finally, taking inspiration from denoising diffusion models [137], the GNN layers at each depth could be replaced with two larger, more expressive models for encoding and decoding across all depths. The resulting models would learn compression and decompression more generally but likely require significantly more training and compression encoding/decoding iterations, as diffusion models do.

Some of the more recent advancements in graph deep learning could also be explored for this problem. Over the last few years, methods using more sophisticated positional and structural encoding have led to significant improvements in graph benchmarks [130, 307]. Local positional encodings, such as node degree, capture a node's relationship to the nodes in its neighbourhood. Global positional encodings, such as eigenvector centrality (a kind of global node degree), represent a node's contribution across the whole graph. Structural encodings, such as participation in graph motifs, indicate the subgraph topological shape around each node. Adding these encodings to node features in the decoder would provide richer graph information for decoding. These encodings could also be incorporated into the tissue microstructure prediction model. For example, global positional encodings would help distinguish between stem villi and chorionic plate nodes, which have very similar cellular node neighbourhoods but are always found in different regions of the slide.

It may also be worth exploring recent graph transformer architectures such as GRIT [129], GraphGPS [130], and Exphormer [131] for capturing long-range node relationships. However, as these methods effectively connect all nodes to all other nodes in a graph, improved autoencoder sampling will be needed to account for the quadratic complexity. GraphGPS achieves linear complexity by decoupling real-edge aggregation from the fully connected transformer [130] but, so far, this has only been scaled to a graph containing 169k nodes [131]. Conversely, it has been shown that adding a virtual node which connects to all other nodes in the graph simulates graph transformers and achieves comparable results [308, 309]. It may be worth exploring virtual nodes first.

7.3 Self-supervised Placental Phenotyping

Self-supervised training has become a popular training method in digital pathology over the last few years. This approach does not need manual training annotations, and models have the freedom to identify their own biological features of importance, but they require large amounts of data to train on. At present, through our clinical international collaborations, we have over 8TB of in-house placenta histology data across a range of pathologies, slide types, and trimesters. Whilst smaller than the data used to train recent foundational digital pathology models, it is likely sufficiently large to train self-supervised models.

For the models themselves, an initial benchmark would be to train a self-supervised GraphSAGE model on cell (or nuclei) graphs using positive and negative neighbourhood sampling for contrastive learning [121]. Here, after neighbourhood aggregation, the model predicts for each node whether a neighbourhood node or a node randomly sampled from elsewhere in the graph is part of its neighbourhood.

This approach results in node features which are grouped in embedding space by neighbourhood similarity that can then be clustered for downstream analysis.

A more recent and popular self-supervised training scheme in histology is the student-teacher self distillation methods of DINO [138] and DINOv2 [139]. Here, two vision transformers [128], student and teacher, output a feature vector for a given image. The student is shown a cropped region of the image which the teacher sees in its entirety. The student is trained to match the outputs of the teacher while the teacher's weights update from an exponentially weighted average of the student's, encouraging the student to learn an efficient representation of the image. DINOv2 is a more computationally efficient implementation trained on a much larger dataset. A DINO model could be trained across all in-house placenta slides of all slide types for region segmentation and instance retrieval of similar biological features. An existing DINO, DINOv2, or other large self-supervised model trained for other organ histology may provide a good pretraining basis for transferring to the placenta [158, 159, 310].

Taking a different approach, denoising diffusion probabilistic models [137] have not yet been explored in digital pathology. Denoising diffusion models use an encoder that gradually adds noise over many steps (usually 1000) and a decoder which learns to reverse this noise by predicting and subtracting it from the image across each step. Denoising diffusion models have seen recent success in image generation [135, 136] and are being explored in the graph space for molecular graphs [311–313]. However, the latent variables of diffusion models do not usually encode representational semantics that can be extracted for downstream analysis [314]. Recently, diffusion autoencoders have introduced an additional semantic encoder to learn such representations [314, 315]. Here, in addition to the encoded noisy image, the decoder receives a semantic embedding from a CNN encoder, which is trained end-to-end from the decoder's outputs. By training from the decoder's image reconstruction, the semantic encoder learns an efficient representation of the original image, which aids in reconstruction. A diffusion autoencoder could be trained across placenta slides either directly on the image data or on cell graphs, where each node is treated as a pixel, and graph compression is handled by the semantic encoder. However, the quality of embeddings extracted from a semantic encoder has not yet been demonstrated for downstream analysis.

To facilitate these proposed methods and encourage future research in digital pathology for the placenta, it would be enormously beneficial to release a large open dataset of placenta histology whole slide images with associated slide characteristics. Currently, there are no open whole slide resources which include the placenta, so AI practitioners are limited to in-house data. The release of such a dataset would additionally allow the placenta to be included in the large multi-organ foundational digital pathology methods.

7.4 Translational Output

Ultimately, developing an automated placental histopathology software tool would provide the most real-world utility for assistive clinical placental pathology and enriching placenta histology research. As an assistive tool, it would need to provide evidence for its outputs by using explainable biology, highlighting regions on the slide, and possibly through a large language model chat interface for discussing the current case. The software should have an easy-to-use interface for non-experts, not require on-site computing resources, and have model predictions that are immediately available by preparing them in the background once slides are digitised.

In addition to a chat interface and slide interaction, the tool can offer instance retrieval by presenting examples of cases with similar slide features and clinical data. The tool can have a built-in self-improvement feedback cycle by flagging cases or slide regions which are unusual to the model. These regions can then be assessed by a perinatal pathologist and added to the training data to improve the model. The tool could assist in pathology report generation by translating from a pathologist's preferred input style, e.g. unstructured prose, into a standardised pathology report [4]. This translation process could also standardise historical pathology reports for future research. Finally, the tool can feature training and examination software for apprentice perinatal pathologists, similar to the interface used in Chapter 5 for comparing perinatal pathologist agreement scores. The goal of such a tool is not to replace perinatal pathologists but to augment their existing workflows, thereby increasing their throughput and freeing up their time for research.

Many of the tool's aforementioned features would also be useful in the research setting. A tool that replicates the same metrics measured by stereology, the current standard practice in placenta histology research, could act as automated stereology across whole slides without manual labelling. Automatically generating whole slide biologically relevant metrics will facilitate future placental histology research across large cohorts.

Looking far into the future, we can imagine a multimodal placental model which combines clinical data, ultrasound imaging taken during pregnancy, macroscopic images of the placenta at delivery, and histology images to offer comprehensive assistive clinical diagnosis. As the efficiency and efficacy of these models and imaging methods improve, placenta investigation may become a standard screening tool for improving maternal and newborn care.

There is no such thing as philosophy-free science; there is only science whose philosophical baggage is taken on board without examination.

— Daniel C. Dennett

8

Ethics Impact Analysis

Contents

8.1 Introduction

With the release and popularity of large language models (LLM) and other generative AI, the general public and funders' appetites for AI solutions have only grown. This includes a desire to use AI for augmenting clinical workflows. However, we must exercise a degree of caution and consider all parties involved when establishing biomedical AI collaborations. AI has the potential to do a lot of good, but if poorly implemented, it could lead to harm.

In this chapter, I will focus on the necessary ethical considerations within the context of biomedical collaboration between AI researchers and developers, AI tools, clinicians, and biomedical researchers. In other words, the creators and maintainers of the tool, the tool itself, and the potential users of the tool. I will draw on examples from my experience developing AI as part of collaborations with perinatal pathologists where relevant. I first lay out the various ethical approvals and other considerations that were made within this project. I will describe examples where AI has been successfully incorporated into clinical workflows and the benefits this has brought. I will then describe a series of potential fail states which need to be accounted for as part of these collaborations.

8.2 Ethical approval of our data and collaborations

The clinical data used throughout this research are from a subset of placentas collected as part of routine clinical pathology investigation from three institutes. The first set is from the University of Tartu (UoT), Estonia; the second set is from the Hadassah Medical Center (HMC), Israel, and the third set is from the Northshore University HealthSystem (NUH), University of Chicago. All samples are fully anonymised to researchers, but data owners at original institutes hold the original data. The use of all pseudo-anonymised samples was approved and the requirement for consent was waived by local ethics committees at each institute (UoT approval 289/T-5 by the Research Ethics Committee of the University of Tartu, HMC approval 0735-18-HMO by the Helsinki Committee at the Hadassah Medical Center, NUH exemption of approval EH23-303 by the institutional review board at the Northshore University HealthSystem given that no clinical data was shared beyond selection for histologically normal term placentas). Data transfer agreements between each original institute and the University of Oxford ensured secure data transfer, storage, and use.

Other ethical considerations taken as part of this research included setting up weekly mentorship meetings with a practising perinatal pathologist as a domain expert. This expert advised on the general direction of the project and feedback on the clinical utility of model outputs. They verified and corrected the manual annotations used to train the model and educated me on placental biology. Having a domain expert involved ensured that any outputs of this research could have further clinical and research utility beyond the scope of this thesis.

For the perinatal pathologist agreement task, a Standard Operating Procedure document was shared to inform pathologists of the nature of the task, what data we would collect on them, and how we would use said data. The interface for performing the task was user-friendly, prioritised pathologists' time, and could be undertaken asynchronously in smaller increments online. On average, it took pathologists one hour to complete the task.

8.3 The potential utility of AI collaborations

There are many examples showing the utility of AI in assisting clinical workflows. In 2016, a collaboration between Moorfields Eye Hospital and DeepMind led to an AI for automated OCT scan analysis [316]. Their initial model matched or surpassed retrospective referral rates compared to clinical experts [317]. Through an ongoing collaboration, this has been expanded to a foundational model for retinal images [318]. In mammography screening for breast cancer, double reading is a labour-intensive protocol where two radiologists assess a case, but it increases cancer detection and lowers recall rates. It was found that replacing one of the readers with an AI system provided similar benefits [319]. Higher diagnostic accuracy was seen in other radiology [320, 321] with explainable AI assistance being particularly useful for non-task experts [322]. In fields where a specialist may not always be available, such as in placenta histopathology when slides are assessed by general surgical pathologists, explainable AI may assist in decreasing the knowledge gap.

AI as a queryable knowledge base may also be useful as a reference source so long as it provides explanations and citations. However, we need to be aware that young clinical trainees are likely to use existing AI solutions to augment their learning. Additional education on how to safely use these systems, including their downsides and caveats, would help mitigate the unreliability of current offerings.

There is a risk, however, that if we are overly cautious when implementing AI solutions, we will not reap the many potential benefits of using this technology.

This caution may partly explain why there has been relatively little application of AI for placenta research. Historically, obtaining ethical approval for releasing pregnancy-related data like placenta histology images has been difficult. The placenta is a rich supply of stem cells for research and therapeutic use, for example, but obtaining approval to use such cells can be challenging [323–325]. With no open-source placenta whole slide datasets, the placenta has been excluded from the large multi-organ digital pathology AI models.

8.4 Risk of overreliance on AI

In high-risk scenarios, such as medical diagnosis, one takes on a degree of responsibility when making a decision. Using third-party systems like AI to assist in the decision-making process can alleviate some of that responsibility, but it is difficult not to overly rely on its confident-sounding outputs. This is especially true if those presented with the outputs are unfamiliar with AI systems and may believe them to be more sophisticated than they are. Care must be taken when designing interfaces to AI interaction to not bias the human users for or against the assisted diagnosis [321, 326, 327].

There is evidence that, when implemented poorly, AI systems can be detrimental to clinical workflows. Assessing the generalisability of commercial AI software for mammography screening on new data, it was found that software-specified thresholds increased recall rates from 5% to 48% [328]. For the same system, a software update led to a three-fold recall rate increase and required version-specific calibration. If such a system had been deployed clinically, it would have been financially disastrous for the institute.

An example of overreliance on AI outside of the medical space is the uptick in LLM usage across academia. Manuscripts with LLM phrases such as “Certainly, I can do . . . for you” are increasing, and LLMs are being mentioned in acknowledgement sections. A large increase in peer review responses containing words such as “intricate”, “commendable”, and “meticulous”, along with other LLM indicators, was seen across top AI conferences [329].

Speaking from personal experience interacting with perinatal pathologists, a perinatal pathologist's job primarily consists of clinical placenta pathology reporting, with a few weeks a year dedicated to placental histopathologic research. There is often a desire to automate parts of the clinical workflow to free up more time for research, especially considering how in-demand perinatal pathologists are. From conversations with perinatal pathologists, a surprisingly common response when we, as AI software developers, make it clear that we are not developing tools to replace their jobs is: "Too bad". AI systems which augment clinical workflows can free up a specialist's time, but if they are seemingly too good, then we may lose the human-in-the-loop.

8.5 AI domain knowledge at deployment

One way to alleviate possible overreliance on AI outputs is to educate the users to model error and overconfidence. AI outputs come with an exaggerated self-specified certainty. For example, an LLM will often respond that it is very certain when you ask it. From a theoretical standpoint, network confidence, the probability assigned to each output class, is often higher than the model's true performance for that class. Models can be highly confident about a certain class of outputs but be wrong in that output most of the time. Aligning class performance and network confidence is an ongoing area of research called calibration [330].

Once a model is deployed, its weights are fixed. The model will always provide the same output from the same input, and its performance in certain areas will not change unless there are explicit changes to the model. Users need to be informed at deployment of the common causes of model error and to be especially cautious in those use cases. A less risky approach would be to avoid offering AI diagnosis directly but output useful, quantifiable metrics with indications for how to verify them from the data. For example, in digital pathology, the model could highlight regions of the slide that informed its output, present similar previous cases, or use a multimodal language model to explain its 'reasoning'.

Finally, collaborators with AI domain knowledge must remain in some capacity to support the deployed AI systems. Without AI domain experts, the model cannot be updated to handle new edge cases or requirements, and any issues with the AI software and software integration cannot be easily resolved by hospital staff or researchers. Once a collaboration has started, it should always exist to some extent.

8.6 The importance of domain experts for AI development

Not only are AI domain experts important in AI deployment, but clinical domain experts are essential in AI development. Clinical experts will direct a project towards clinical utility, they can highlight any issues with the data, they will provide sanity checks for model outputs, and they are best placed to interpret the results within a clinical context.

To take an example from the consumer tech space, when Fitbit introduced a period tracker to its calendar, it did not allow women to input periods longer than 11 days [331]. In this case, the domain experts that they lacked were women. There is also evidence that when AI developers only focus on performance in benchmark datasets (which are made by AI developers for other AI developers), the models will be unable to transfer to real-world data [332, 333].

Speaking from personal experience, I did not have direct access to a domain expert for the first year of my doctorate. I educated myself as best I could from online resources, but there were gaps in my understanding. For example, I had misunderstood that there was a difference between 'immature intermedial villi' and 'mature intermediate villi'. After the first year, we collaborated with another perinatal pathologist, and I had short weekly mentorship meetings with this pathologist for the rest of my degree. This mentorship was invaluable to my understanding, to the quality of manual data annotations, and to direct the research in a way that it could have clinical utility.

8.7 The unnamed stakeholder: the patient

The final ethical considerations for biomedical AI collaborations that I will cover are the patients who indirectly benefit from but may also be harmed by these collaborations. Patients are often not explicitly aware of how hospitals may use their data which was collected as part of standard clinical investigation. It is generally understood that such data may be used to train clinical staff within the hospital and, with the correct ethical approvals, for research by the hospital or associated university. However, patients may not be cognizant of the extent to which their data can be used in research. With the advent of AI in research, it is an open question whether patients would (or could) object to their data being used to train AI systems. We must also ensure that such AI systems do not have the capability to output their training data, thereby breaking patient confidentiality. Regarding diagnosis, we need to establish whether patients will be informed that an assistive AI has been used for their diagnosis and if they can opt out in advance. Finally, when considering placenta and pregnancy-related research and diagnoses, we must be aware that there are two (or more) patients involved.

References

[1] Ona Marie Faye-Petersen and Linda M Ernst. “Maternal Floor Infarction and Massive Perivillous Fibrin Deposition”. en. In: *Surg. Pathol. Clin.* 6.1 (Mar. 2013), pp. 101–114.

[2] Raymond W Redline. “Extending the Spectrum of Massive Perivillous Fibrin Deposition (Maternal Floor Infarction)”. en. In: *Pediatr. Dev. Pathol.* 24.1 (2021), pp. 10–11.

[3] Raymond W Redline et al. “Placental pathology is necessary to understand common pregnancy complications and achieve an improved taxonomy of obstetrical disease”. English. In: *Am. J. Obstet. Gynecol.* 228.2 (Feb. 2023), pp. 187–202.

[4] Gitta Turowski et al. “The structure and utility of the placental pathology report”. en. In: *APMIS* 126.7 (2018), pp. 638–646.

[5] A E P Heazell and E A Martindale. “Can post-mortem examination of the placenta help determine the cause of stillbirth?” en. In: *J. Obstet. Gynaecol.* 29.3 (Apr. 2009), pp. 225–228.

[6] Carmen Scalise et al. “The Importance of Post-Mortem Investigations in Stillbirths: Case Studies and a Review of the Literature”. en. In: *Int. J. Environ. Res. Public Health* 19.14 (July 2022).

[7] Beatrix B Thompson, Parker H Holzer, and Harvey J Kliman. “Placental Pathology Findings in Unexplained Pregnancy Losses”. en. In: *Reprod. Sci.* (Sept. 2023).

[8] Gitta Turowski, Susan Arbuckle, and W Tony Parks. “The Placental Pathology Report”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 371–377.

[9] Sunitha C Suresh et al. “A comprehensive analysis of the association between placental pathology and recurrent preterm birth”. en. In: *Am. J. Obstet. Gynecol.* 227.6 (Dec. 2022), 887.e1–887.e15.

[10] Sophie Stenton et al. “SARS-COV2 placentitis and pregnancy outcome: A multicentre experience during the Alpha and early Delta waves of coronavirus pandemic in England”. English. In: *eClinicalMedicine* 47 (May 2022).

[11] David A Schwartz, Sarah B Mulkey, and Drucilla J Roberts. “SARS-CoV-2 placentitis, stillbirth, and maternal COVID-19 vaccination: clinical–pathologic correlations”. In: *Am. J. Obstet. Gynecol.* 228.3 (2023), pp. 261–269.

[12] Evelyn Annegret Huhn et al. “Fetal Programming”. In: *The Placenta: Basics and Clinical Significance.* Ed. by Berthold Huppertz and Ekkehard Schleußner. Berlin, Heidelberg: Springer Berlin Heidelberg, 2023, pp. 295–315.

[13] Graham J Burton, Abigail L Fowden, and Kent L Thornburg. “Placental Origins of Chronic Disease”. eng. In: *Physiol. Rev.* 96.4 (Oct. 2016), pp. 1509–1565.

[14] Salvatore Lacagnina. “The Developmental Origins of Health and Disease (DOHaD)”. en. In: *Am. J. Lifestyle Med.* 14.1 (2020), pp. 47–50.

[15] Eun Jin Kwon and Young Ju Kim. “What is fetal programming?: a lifetime health is under the control of in utero health”. In: *Obstet Gynecol Sci* 60.6 (Nov. 2017), pp. 506–519.

[16] D J P Barker and K L Thornburg. “Placental programming of chronic diseases, cancer and lifespan: a review”. eng. In: *Placenta* 34.10 (Oct. 2013), pp. 841–845.

[17] Carlos Escudero et al. “Brain Vascular Dysfunction in Mothers and Their Children Exposed to Preeclampsia”. en. In: *Hypertension* 80.2 (Feb. 2023), pp. 242–256.

[18] María Paternina-Die et al. “Women’s neuroplasticity during gestation, childbirth and postpartum”. en. In: *Nat. Neurosci.* 27.2 (Feb. 2024), pp. 319–327.

[19] Cynthia G Kaplan and Rebecca N Baergen. “Approach to Histologic Examination of the Placenta”. In: *Benirschke’s Pathology of the Human Placenta.* Ed. by Rebecca N Baergen, Graham J Burton, and Cynthia G Kaplan. Cham: Springer International Publishing, 2022, pp. 223–226.

[20] Tim H H Coorens et al. “Inherent mosaicism and extensive mutation of human placentas”. en. In: *Nature* 592.7852 (Apr. 2021), pp. 80–85.

[21] Roberto Romero et al. “The Frequency and Type of Placental Histologic Findings in Term Pregnancies with Normal Outcome”. In: *J. Perinat. Med.* 46.6 (Aug. 2018), pp. 613–630.

[22] Drucilla J Roberts et al. “Criteria for placental examination for obstetric and neonatal providers1”. en. In: *Am. J. Obstet. Gynecol.* (Dec. 2022).

[23] Amy Heerema-McKenney. “Placental Development with Expected Normal Gross and Microscopic Findings”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 9–29.

[24] Mudher Al-Adnani et al. ““Delayed Villous Maturation” in Placental Reporting: Concordance among Consultant Pediatric Pathologists at a Single Specialist Center”. en. In: *Pediatr. Dev. Pathol.* 18.5 (Sept. 2015), pp. 375–379.

[25] Jason W Custer et al. “Diagnostic errors in the pediatric and neonatal ICU: a systematic review”. eng. In: *Pediatr. Crit. Care Med.* 16.1 (Jan. 2015), pp. 29–36.

[26] Anika Mukherjee et al. “The Placental Distal Villous Hypoplasia Pattern: Interobserver Agreement and Automated Fractal Dimension as an Objective Metric”. en. In: *Pediatr. Dev. Pathol.* 19.1 (2016), pp. 31–36.

[27] Jennifer K Straughen et al. “Methods to decrease variability in histological scoring in placentas from a cohort of preterm infants”. en. In: *BMJ Open* 7.3 (Mar. 2017), e013877.

[28] Chen-Chih J Sun et al. “Discrepancy in pathologic diagnosis of placental lesions”. eng. In: *Arch. Pathol. Lab. Med.* 126.6 (June 2002), pp. 706–709.

[29] Linda M Ernst et al. "Comparison of Placental Pathology Reports From Spontaneous Preterm Births Finalized by General Surgical Pathologists Versus Perinatal Pathologist: A Call to Action". en-US. In: *Am. J. Surg. Pathol.* 47.10 (Oct. 2023), p. 1116.

[30] Imelda Odibo et al. "Pathologic examination of the placenta and its clinical utility: a survey of obstetrics and gynecology providers". en. In: *J. Matern. Fetal. Neonatal Med.* 29.2 (2016), pp. 197–201.

[31] Beata Hargitai, Tamas Marton, and Amy Heerema-McKenney. "Indications for Examining the Placenta". en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 31–37.

[32] The Royal College of Pathologists. *Become a paediatric and perinatal pathologist.* `https://www.rcpath.org/discover-pathology/careers-in-pathology/careers-in-medicine/become-a-paediatric-and-perinatal-pathologist.html`. Accessed: 2021-11-21.

[33] Megan Munro and Meg Lyons. *Births in England and Wales - Office for National Statistics.* en. `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2022refreshedpopulations`. Accessed: 2024-4-16. Feb. 2024.

[34] Brock E Polnaszek, Steven L Clark, and Dwight J Rouse. "Pathologic Assessment of the Placenta: Evidence Compared With Tradition". en-US. In: *Obstetrics & Gynecology* 139.4 (Apr. 2022), pp. 660–667.

[35] William M Curtin et al. "Pathologic examination of the placenta and observed practice". en. In: *Obstet. Gynecol.* 109.1 (Jan. 2007), pp. 35–41.

[36] Aysha Mubeen and Raafat Makary. "Pathologic examination of placenta: a study on 500 live births to assess conformity to College of American Pathologists (CAP) guidelines and clinicopathologic correlation". en. In: *J. Matern. Fetal. Neonatal Med.* 35.20 (Oct. 2022), pp. 3867–3871.

[37] Soo Downe et al. "Post-mortem examination after stillbirth: views of UK-based practitioners". en. In: *Eur. J. Obstet. Gynecol. Reprod. Biol.* 162.1 (May 2012), pp. 33–37.

[38] Amy Watts, Charlotte Hill, and Zoe Jones. *Child and infant mortality in England and Wales - Office for National Statistics.* en. `https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/childhoodinfantandperinatalmortalityinenglandandwales/2021`. Accessed: 2024-4-16. Mar. 2023.

[39] Sands. *Bereaved Parents Experience of Care Report 2023.* Tech. rep.

[40] A Garrod et al. "Duration and method of tissue storage alters placental morphology – Implications for clinical and research practice". en. In: *Placenta* 34.11 (Nov. 2013), pp. 1116–1119.

[41] Rebecca N Baergen. "Examination of the Placenta". In: *Benirschke's Pathology of the Human Placenta.* Ed. by Rebecca N Baergen, Graham J Burton, and Cynthia G Kaplan. Cham: Springer International Publishing, 2022, pp. 1–9.

[42] S J Gordijn et al. "Histopathological examination of the placenta: key issues for pathologists and obstetricians". eng. In: *Pathology* 40.2 (Feb. 2008), pp. 176–179.

[43] B Hargitai, T Marton, and P M Cox. "BEST PRACTICE NO 178 : Examination of the human placenta". en. In: *J. Clin. Pathol.* 57.8 (Aug. 2004), pp. 785–792.

[44] Nikos Papadogiannakis. "How can we ensure that placental pathology plays a greater role in treating neonatal morbidity?" en. In: *Acta Paediatr.* 109.8 (2020), pp. 1514–1515.

[45] Natalie D Mercuri and Brian J Cox. "The need for more research into reproductive health and disease". In: *Elife* 11 (Dec. 2022). Ed. by Peter Rodgers, Marleen van Gelder, and James Roberts, e75061.

[46] T M Mayhew. "Stereology and the placenta: where's the point? – a review". en. In: *Placenta* 27 Suppl A (Apr. 2006), S17–25.

[47] Hannah Ee Juen Yong et al. "Integrated Placental Modelling of Histology with Gene Expression to Identify Functional Impact on Fetal Growth". en. In: *Cells* 12.7 (Apr. 2023).

[48] Amani M Abdalla, Muddathir D Tingari, and Mohamed A Abdalla. "Histomorphometric parameters of normal full term placenta of Sudanese women". en. In: *Heliyon* 2.7 (July 2016), e00135.

[49] Andrew H Song et al. "Artificial intelligence for digital and computational pathology". en. In: *Nature Reviews Bioengineering* 1.12 (Oct. 2023), pp. 930–949.

[50] Michael Ferlaino et al. "Towards Deep Cellular Phenotyping in Placental Histology". In: *arXiv:1804.03270 [cs]* (May 2018).

[51] Pooya Mobadersany, Lee A D Cooper, and Jeffery A Goldstein. "GestAltNet: aggregation and attention to improve deep learning of gestational age from placental whole-slide images". en. In: *Lab. Invest.* 101.7 (July 2021), pp. 942–951.

[52] Daniel Clymer et al. "Decidual Vasculopathy Identification in Whole Slide Images Using Multiresolution Hierarchical Convolutional Neural Networks". English. In: *Am. J. Pathol.* 190.10 (Oct. 2020), pp. 2111–2122.

[53] Jeffery A Goldstein et al. "Machine learning classification of placental villous infarction, perivillous fibrin deposition, and intervillous thrombus". In: *Placenta* 135 (Apr. 2023), pp. 43–50.

[54] Purvasha Patnaik et al. "Automated detection of microscopic placental features indicative of maternal vascular malperfusion using machine learning". en. In: *Placenta* 145 (Jan. 2024), pp. 19–26.

[55] Leticia Reyes and Thaddeus G Golos. "Hofbauer Cells: Their Role in Healthy and Complicated Pregnancy". In: *Front. Immunol.* 9 (2018), p. 2628.

[56] Berthold Huppertz. "Placental Development with Histological Aspects". In: *The Placenta: Basics and Clinical Significance.* Ed. by Berthold Huppertz and Ekkehard Schleußner. Berlin, Heidelberg: Springer Berlin Heidelberg, 2023, pp. 1–27.

[57] Edward B Chuong. "The placenta goes viral: Retroviruses control gene expression in pregnancy". In: *PLoS Biol.* 16.10 (Oct. 2018), e3000028.

[58] S Mi et al. “Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis”. en. In: *Nature* 403.6771 (Feb. 2000), pp. 785–789.

[59] Kazuhiko Imakawa, So Nakagawa, and Takayuki Miyazawa. “Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution”. en. In: *Genes Cells* 20.10 (Oct. 2015), pp. 771–788.

[60] Anthony M Carter. “Unique Aspects of Human Placentation”. en. In: *Int. J. Mol. Sci.* 22.15 (July 2021).

[61] Linda M Ernst and Chrystalle Katte Carreon. “Placenta”. en. In: *Color Atlas of Human Fetal and Neonatal Histology*. Ed. by Linda M Ernst et al. Cham: Springer International Publishing, 2019, pp. 399–424.

[62] Yashvardhan Jain et al. “Segmenting functional tissue units across human organs using community-driven development of generalizable machine learning algorithms”. en. In: *Nat. Commun.* 14.1 (Aug. 2023), p. 4656.

[63] Sharmistha Biswas, Samit K Ghosh, and Shakuntala Chhabra. “Surface area of chorionic villi of placentas: an index of intrauterine growth restriction of fetuses”. en. In: *J. Obstet. Gynaecol. Res.* 34.4 (Aug. 2008), pp. 487–493.

[64] Graham J Burton. “Architecture of the Villous Trees”. In: *Benirschke’s Pathology of the Human Placenta*. Ed. by Rebecca N Baergen, Graham J Burton, and Cynthia G Kaplan. Cham: Springer International Publishing, 2022, pp. 111–141.

[65] Gitta Turowski and Martin Vogel. “Re-view and view on maturation disorders in the placenta”. en. In: *APMIS* 126.7 (2018), pp. 602–612.

[66] Gerit Moser et al. “Human trophoblast invasion: new and unexpected routes and functions”. en. In: *Histochem. Cell Biol.* 150.4 (Oct. 2018), pp. 361–370.

[67] Graham J Burton and Eric Jauniaux. “The Chorionic and Basal Plates”. In: *Benirschke’s Pathology of the Human Placenta*. Ed. by Rebecca N Baergen, Graham J Burton, and Cynthia G Kaplan. Cham: Springer International Publishing, 2022, pp. 205–221.

[68] C D Ockleford. “The allo-epi-endothelial lining of the intervillous space”. en. In: *Placenta* 31.12 (Dec. 2010), pp. 1035–1042.

[69] Ian A Hatton et al. “The human cell count and size distribution”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 120.39 (Sept. 2023), e2303077120.

[70] Yuping Wang and Shuang Zhao. *Vascular Biology of the Placenta*. eng. Integrated Systems Physiology: from Molecules to Function to Disease. San Rafael (CA): Morgan & Claypool Life Sciences, 2010.

[71] Harold Fox and Neil J Sebire. “2 - THE DEVELOPMENT AND STRUCTURE OF THE PLACENTA”. In: *Pathology of the Placenta (Third Edition)*. Ed. by Harold Fox and Neil J Sebire. Edinburgh: W.B. Saunders, Jan. 2007, pp. 17–56.

[72] W Tony Parks. “Increased Syncytial Knot Formation”. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 131–137.

[73] Yawei Liu et al. “Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta”. en. In: *Cell Res.* 28.8 (Aug. 2018), pp. 819–832.

[74] Mana Parast. “Normal Development”. In: *Placental and Gestational Pathology.* Ed. by Raymond W Redline, Theonia K Boyd, and Drucilla J Editors Roberts. Diagnostic Pediatric Pathology. Cambridge University Press, 2017, pp. 1–8.

[75] Harold Fox and Neil J Sebire. “3 - PHYSIOLOGY OF THE PLACENTA”. In: *Pathology of the Placenta (Third Edition).* Ed. by Harold Fox and Neil J Sebire. Edinburgh: W.B. Saunders, Jan. 2007, pp. 57–67.

[76] T Yee Khong. “Persistence of Cytotrophoblast”. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 139–141.

[77] T Yee Khong et al. “Sampling and Definitions of Placental Lesions: Amsterdam Placental Workshop Group Consensus Statement”. en. In: *Arch. Pathol. Lab. Med.* 140.7 (July 2016), pp. 698–713.

[78] Kurt Benirschke, Graham J Burton, and Rebecca N Baergen. *Pathology of the Human Placenta.* Springer Berlin Heidelberg, 2012.

[79] Kristina Loukeris, Raanan Sela, and Rebecca N Baergen. “Syncytial knots as a reflection of placental maturity: reference values for 20 to 40 weeks’ gestational age”. eng. In: *Pediatr. Dev. Pathol.* 13.4 (July 2010), pp. 305–309.

[80] Theonia K Boyd, Drucilla J Roberts, and Amy Heerema-McKenney. “Fetal Vascular Malperfusion”. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 173–182.

[81] Gitta Turowski, Eoghan E Mooney, and Irene B Scheimberg. “Maturity Anomalies”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 115–121.

[82] Omar Farah et al. “Trophoblast lineage-specific differentiation and associated alterations in preeclampsia and fetal growth restriction”. en. In: *Placenta* 102 (Dec. 2020), pp. 4–9.

[83] T Yee Khong. “Extravillous Trophoblast Cyst”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 97–99.

[84] Brendan Fitzgerald. “Infarction”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 57–65.

[85] Olav Lapaire et al. “Fetal CellsFetal cells and Cell-Free Nucleic Acids in Maternal Blood: Genetic and Immunological Aspects”. In: *The Placenta: Basics and Clinical Significance.* Ed. by Berthold Huppertz and Ekkehard Schleußner. Berlin, Heidelberg: Springer Berlin Heidelberg, 2023, pp. 317–332.

[86] Kurt Benirschke, Graham J Burton, and Rebecca N Baergen. “Architecture of Normal Villous Trees”. en. In: *Pathology of the Human Placenta.* Ed. by Kurt Benirschke, Graham J Burton, and Rebecca N Baergen. Berlin, Heidelberg: Springer, 2012, pp. 101–144.

[87] M R Riddell et al. “The characterization of fibrocyte-like cells: a novel fibroblastic cell of the placenta”. en. In: *Placenta* 33.3 (Mar. 2012), pp. 143–150.

[88] Dong-Bao Chen and Jing Zheng. “Regulation of placental angiogenesis”. en. In: *Microcirculation* 21.1 (Jan. 2014), pp. 15–25.

[89] Zhonghua Tang et al. “Placental Hofbauer cells and complications of pregnancy”. en. In: *Ann. N. Y. Acad. Sci.* 1221 (Mar. 2011), pp. 103–108.

[90] Charalampos Grigoriadis et al. “Hofbauer cells morphology and density in placentas from normal and pathological gestations”. eng. In: *Rev. Bras. Ginecol. Obstet.* 35.9 (Sept. 2013), pp. 407–412.

[91] Hidetaka Okada, Tomoko Tsuzuki, and Hiromi Murata. “Decidualization of the human endometrium”. en. In: *Reprod. Med. Biol.* 17.3 (July 2018), pp. 220–227.

[92] Shipra Sharma, Geeta Godbole, and Deepak Modi. “Decidual Control of Trophoblast Invasion”. en. In: *Am. J. Reprod. Immunol.* 75.3 (Mar. 2016), pp. 341–350.

[93] Linda M Ernst, Suzanne M Jacques, and Faisal Qureshi. “Basal Plate Myometrial Fibres”. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 237–241.

[94] Tamara Garrido-Gomez et al. “Defective decidualization during and after severe preeclampsia reveals a possible maternal contribution to the etiology”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 114.40 (Oct. 2017), E8468–E8477.

[95] Phillip Cox, Marta C Cohen, and Irene B Scheimberg. “Acute Chorioamnionitis”. en. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 103–107.

[96] Rodrigo Vega-Sanchez et al. “Placental blood leukocytes are functional and phenotypically different than peripheral leukocytes during human labor”. en. In: *J. Reprod. Immunol.* 84.1 (Jan. 2010), pp. 100–110.

[97] Marta C Cohen and Theonia K Boyd. “Presence of Nucleated Red Blood Cells”. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 183–186.

[98] Katarzyna Pikora et al. “Diagnostic Value and Prognostic Significance of Nucleated Red Blood Cells (NRBCs) in Selected Medical Conditions”. en. In: *Cells* 12.14 (July 2023).

[99] C Langston et al. “Practice guideline for examination of the placenta: developed by the Placental Pathology Practice Guideline Development Task Force of the College of American Pathologists”. eng. In: *Arch. Pathol. Lab. Med.* 121.5 (May 1997), pp. 449–476.

[100] Clair Evans and Phillip Cox. *Tissue pathway for histopathological examination of the placenta.* 2019.

[101] S M Powsner, J Costa, and R J Homer. “Clinicians are from Mars and pathologists are from Venus”. eng. In: *Arch. Pathol. Lab. Med.* 124.7 (July 2000), pp. 1040–1046.

[102] Anna Lieschke et al. “Uncertainty over implications of placental histopathological findings: A survey of Australian and New Zealand neonatologists”. en. In: *J. Paediatr. Child Health* 56.2 (2020), pp. 259–264.

[103] Alexa A Freedman et al. “Formulating a Meaningful and Comprehensive Placental Phenotypic Classification”. eng. In: *Pediatr. Dev. Pathol.* 24.4 (July 2021), pp. 337–350.

[104] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. “Learning representations by back-propagating errors”. en. In: *Nature* 323.6088 (Oct. 1986), pp. 533–536.

[105] F Rosenblatt. “The perceptron: a probabilistic model for information storage and organization in the brain”. en. In: *Psychol. Rev.* 65.6 (Nov. 1958), pp. 386–408.

[106] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis and Machine Vision.* Springer US.

[107] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. “ImageNet classification with deep convolutional neural networks”. In: *Commun. ACM* 60.6 (May 2017), pp. 84–90.

[108] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition.* June 2009, pp. 248–255.

[109] K Fukushima. “Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position”. en. In: *Biol. Cybern.* 36.4 (1980), pp. 193–202.

[110] Y LeCun et al. “Backpropagation Applied to Handwritten Zip Code Recognition”. In: *Neural Comput.* 1.4 (Dec. 1989), pp. 541–551.

[111] Y Lecun et al. “Gradient-based learning applied to document recognition”. In: *Proc. IEEE* 86.11 (Nov. 1998), pp. 2278–2324.

[112] Karen Simonyan and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: (Sept. 2014). arXiv: `1409.1556 [cs.CV]`.

[113] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *arXiv:1512.03385 [cs]* (Dec. 2015).

[114] Tsung-Yi Lin et al. “Focal Loss for Dense Object Detection”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 42.2 (Feb. 2020), pp. 318–327.

[115] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors”. In: (July 2022). arXiv: `2207.02696 [cs.CV]`.

[116] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* Springer International Publishing, 2015, pp. 234–241.

[117] Kaiming He et al. “Mask R-CNN”. In: (Mar. 2017). arXiv: `1703.06870 [cs.CV]`.

[118] Thomas N Kipf and Max Welling. *Semi-Supervised Classification with Graph Convolutional Networks.* Feb. 2017.

[119] Rex Ying et al. “Graph Convolutional Neural Networks for Web-Scale Recommender Systems”. In: (June 2018). arXiv: `1806.01973 [cs.IR]`.

[120] Hang Zhou et al. “Graph Neural Network for Protein-Protein Interaction Prediction: A Comparative Study”. en. In: *Molecules* 27.18 (Sept. 2022).

[121] William L Hamilton, Rex Ying, and Jure Leskovec. “Inductive Representation Learning on Large Graphs”. en. In: June 2017.

[122] Milad Besharatifard and Fatemeh Vafaee. “A review on graph neural networks for predicting synergistic drug combinations”. In: *Artificial Intelligence Review* 57.3 (Feb. 2024), p. 49.

[123] Wei-Lin Chiang et al. “Cluster-GCN: An Efficient Algorithm for Training Deep and Large Graph Convolutional Networks”. en. In: *KDD’19*. May 2019.

[124] Hanqing Zeng et al. “GraphSAINT: Graph Sampling Based Inductive Learning Method”. In: *arXiv:1907.04931 [cs, stat]*. Feb. 2020.

[125] Weihua Hu et al. “Open Graph Benchmark: Datasets for Machine Learning on Graphs”. In: *arXiv:2005.00687 [cs, stat]* (Feb. 2021).

[126] Weihua Hu et al. “OGB-LSC: A Large-Scale Challenge for Machine Learning on Graphs”. In: (Mar. 2021). arXiv: `2103.09430 [cs.LG]`.

[127] Xin Liu et al. “Survey on Graph Neural Network Acceleration: An Algorithmic Perspective”. In: (Feb. 2022). arXiv: `2202.04822 [cs.LG]`.

[128] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: (Oct. 2020). arXiv: `2010.11929 [cs.CV]`.

[129] Liheng Ma et al. “Graph Inductive Biases in Transformers without Message Passing”. In: (May 2023). arXiv: `2305.17589 [cs.LG]`.

[130] Ladislav Rampášek et al. “Recipe for a General, Powerful, Scalable Graph Transformer”. In: (May 2022). arXiv: `2205.12454 [cs.LG]`.

[131] Hamed Shirzad et al. “Exphormer: Sparse Transformers for Graphs”. In: (Mar. 2023). arXiv: `2303.06147 [cs.LG]`.

[132] Ashish Vaswani et al. “Attention Is All You Need”. In: (June 2017). arXiv: `1706.03762 [cs.CL]`.

[133] Samuel L Smith et al. “ConvNets Match Vision Transformers at Scale”. In: (Oct. 2023). arXiv: `2310.16764 [cs.CV]`.

[134] G E Hinton and R R Salakhutdinov. “Reducing the dimensionality of data with neural networks”. en. In: *Science* 313.5786 (July 2006), pp. 504–507.

[135] Robin Rombach et al. “High-Resolution Image Synthesis with Latent Diffusion Models”. In: (Dec. 2021). arXiv: `2112.10752 [cs.CV]`.

[136] Aditya Ramesh et al. “Hierarchical Text-Conditional Image Generation with CLIP Latents”. In: (Apr. 2022). arXiv: `2204.06125 [cs.CV]`.

[137] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. In: (June 2020). arXiv: `2006.11239 [cs.LG]`.

[138] Mathilde Caron et al. “Emerging Properties in Self-Supervised Vision Transformers”. In: (Apr. 2021). arXiv: `2104.14294 [cs.CV]`.

[139] Maxime Oquab et al. “DINOv2: Learning Robust Visual Features without Supervision”. In: (Apr. 2023). arXiv: `2304.07193 [cs.CV]`.

[140] Tom B Brown et al. “Language Models are Few-Shot Learners”. In: (May 2020). arXiv: `2005.14165 [cs.CL]`.

[141] Rishi Bommasani et al. “On the Opportunities and Risks of Foundation Models”. In: (Aug. 2021). arXiv: `2108.07258 [cs.LG]`.

[142] OpenAI et al. “GPT-4 Technical Report”. In: (Mar. 2023). arXiv: `2303.08774 [cs.CL]`.

[143] Alexander Kirillov et al. “Segment Anything”. In: (Apr. 2023). arXiv: `2304.02643 [cs.CV]`.

[144] Peter Bankhead. *QuPath: Open source software for digital pathology image analysis | Scientific Reports.* `https://www.nature.com/articles/s41598-017-17204-5`. Accessed: 2021-11-9. 2017.

[145] Babak Ehteshami Bejnordi et al. “Quantitative analysis of stain variability in histology slides and an algorithm for standardization”. In: *Medical Imaging 2014: Digital Pathology.* Vol. 9041. SPIE, Mar. 2014, pp. 45–51.

[146] Elizabeth A Chlipala et al. “Impact of Preanalytical Factors During Histology Processing on Section Suitability for Digital Image Analysis”. en. In: *Toxicol. Pathol.* 49.4 (June 2021), pp. 755–772.

[147] Geoff Dougherty. *Digital Image Processing for Medical Applications.* Cambridge, England: Cambridge University Press, 2009.

[148] David Tellez et al. “Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology”. en. In: *Med. Image Anal.* 58 (Dec. 2019), p. 101544.

[149] Khrystyna Faryna, Jeroen van der Laak, and Geert Litjens. “Automatic data augmentation to improve generalization of deep learning in H&E stained histopathology”. en. In: *Comput. Biol. Med.* 170 (Mar. 2024), p. 108018.

[150] Maryam Haghighat et al. “Automated quality assessment of large digitised histology cohorts by artificial intelligence”. en. In: *Sci. Rep.* 12.1 (Mar. 2022), p. 5002.

[151] Neel Kanwal et al. “Are you sure it’s an artifact? Artifact detection and uncertainty quantification in histological images”. In: *Comput. Med. Imaging Graph.* 112 (Mar. 2024), p. 102321.

[152] Le Hou et al. “Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* June 2016, pp. 2424–2433.

[153] Angel Cruz-Roa et al. “Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks”. en. In: *Medical Imaging 2014: Digital Pathology.* Vol. 9041. SPIE, Mar. 2014, p. 904103.

[154] Simon Graham et al. “Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images”. In: *Med. Image Anal.* 58 (Dec. 2019), p. 101563.

[155] David Ahmedt-Aristizabal et al. “A survey on graph-based deep learning for computational histopathology”. en. In: *Comput. Med. Imaging Graph.* 95 (Jan. 2022), p. 102027.

[156] Ming Y Lu et al. “A Foundational Multimodal Vision Language AI Assistant for Human Pathology”. In: (Dec. 2023). arXiv: `2312.07814 [cs.CV]`.

[157] Jun Ma et al. “Segment anything in medical images”. en. In: *Nat. Commun.* 15.1 (Jan. 2024), p. 654.

[158] Richard J Chen et al. “Towards a general-purpose foundation model for computational pathology”. en. In: *Nat. Med.* (Mar. 2024), pp. 1–13.

[159] Ming Y Lu et al. “A visual-language foundation model for computational pathology”. en. In: *Nat. Med.* (Mar. 2024), pp. 1–12.

[160] Xiyue Wang et al. “Transformer-based unsupervised contrastive learning for histopathological image classification”. en. In: *Med. Image Anal.* 81 (Oct. 2022), p. 102559.

[161] Zhi Huang et al. “A visual-language foundation model for pathology image analysis using medical Twitter”. en. In: *Nat. Med.* 29.9 (Sept. 2023), pp. 2307–2316.

[162] Rayan Krishnan, Pranav Rajpurkar, and Eric J Topol. “Self-supervised learning in medicine and healthcare”. en. In: *Nat Biomed Eng* 6.12 (Dec. 2022), pp. 1346–1352.

[163] Fahad Shamshad et al. “Transformers in medical imaging: A survey”. en. In: *Med. Image Anal.* 88 (Aug. 2023), p. 102802.

[164] Richard J Chen et al. “Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning”. en. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* New Orleans, LA, USA: IEEE, 2022, pp. 16123–16134.

[165] Fabian Hörst et al. “CellViT: Vision Transformers for precise cell segmentation and classification”. In: *Med. Image Anal.* (Mar. 2024), p. 103143.

[166] Jevgenij Gamper et al. “PanNuke: An Open Pan-Cancer Histology Dataset for Nuclei Instance Segmentation and Classification”. In: *Digital Pathology.* Springer International Publishing, 2019, pp. 11–19.

[167] Jevgenij Gamper et al. *PanNuke Dataset Extension, Insights and Baselines.* Apr. 2020.

[168] S Graham et al. “Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification”. In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (Aug. 2021), pp. 684–693.

[169] Dequan Wang et al. “A Real-world Dataset and Benchmark For Foundation Model Adaptation in Medical Image Classification”. en. In: *Sci Data* 10.1 (Sept. 2023), p. 574.

[170] Peter Bandi et al. “From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge”. en. In: *IEEE Trans. Med. Imaging* 38.2 (Feb. 2019), pp. 550–560.

[171] Guilherme Aresta et al. “BACH: Grand challenge on breast cancer histology images”. en. In: *Med. Image Anal.* 56 (Aug. 2019), pp. 122–139.

[172] Korsuk Sirinukunwattana et al. “Gland segmentation in colon histology images: The glas challenge contest”. en. In: *Med. Image Anal.* 35 (Jan. 2017), pp. 489–502.

[173] Jeongun Ryu et al. “OCELOT: Overlapped Cell on Tissue Dataset for Histopathology”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2023, pp. 23902–23912.

[174] Andrew H Fischer et al. “Hematoxylin and eosin staining of tissue and cell sections”. eng. In: *CSH Protoc.* 2008 (May 2008), db.prot4986.

[175] Claudia Vanea et al. “Mapping cell-to-tissue graphs across human placenta histology whole slide images using deep learning with HAPPY”. en. In: *Nat. Commun.* 15.1 (Mar. 2024), pp. 1–16.

[176] Nobuyuki Otsu. “A Threshold Selection Method from Gray-Level Histograms”. In: *IEEE Trans. Syst. Man Cybern.* 9.1 (Jan. 1979), pp. 62–66.

[177] F Meyer and S Beucher. “Morphological segmentation”. In: *J. Vis. Commun. Image Represent.* 1.1 (Sept. 1990), pp. 21–46.

[178] Jean-Romain Dalle et al. “Automatic breast cancer grading of histopathological images”. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, Aug. 2008, pp. 3052–3055.

[179] Mark D Zarella et al. “An alternative reference space for H&E color normalization”. en. In: *PLoS One* 12.3 (Mar. 2017), e0174489.

[180] Ruchika Verma et al. “MoNuSAC2020: A Multi-Organ Nuclei Segmentation and Classification Challenge”. In: *IEEE Trans. Med. Imaging* 40.12 (Dec. 2021), pp. 3413–3423.

[181] Tan N N Doan et al. “SONNET: A Self-Guided Ordinal Regression Neural Network for Segmentation and Classification of Nuclei in Large-Scale Multi-Tissue Histology Images”. In: *IEEE Journal of Biomedical and Health Informatics* 26.7 (July 2022), pp. 3218–3228.

[182] Quoc Dang Vu et al. “Methods for Segmentation and Classification of Digital Microscopy Tissue Images”. en. In: *Front Bioeng Biotechnol* 7 (Apr. 2019), p. 53.

[183] K Martinez and J Cupitt. “VIPS - a highly tuned image processing software architecture”. In: *IEEE International Conference on Image Processing 2005.* Vol. 2. Sept. 2005, pp. II–574.

[184] Melissa Linkert et al. “Metadata matters: access to image data in the real world”. In: *J. Cell Biol.* 189.5 (May 2010), pp. 777–782.

[185] Adam Goode et al. “OpenSlide: A vendor-neutral software foundation for digital pathology”. en. In: *J. Pathol. Inform.* 4.1 (Jan. 2013), p. 27.

[186] Jon Louis Bentley. “Multidimensional binary search trees used for associative searching”. In: *Commun. ACM* 18.9 (Sept. 1975), pp. 509–517.

[187] Tsung-Yi Lin et al. “Microsoft COCO: Common Objects in Context”. en. In: *Computer Vision – ECCV 2014.* Ed. by David Fleet et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014, pp. 740–755.

[188] Diederik P Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. en. In: Dec. 2014.

[189] Alexander Buslaev et al. “Albumentations: Fast and Flexible Image Augmentations”. en. In: *Information* 11.2 (Feb. 2020), p. 125.

[190] A C Ruifrok and D A Johnston. “Quantification of histochemical staining by color deconvolution”. eng. In: *Anal. Quant. Cytol. Histol.* 23.4 (Aug. 2001), pp. 291–299.

[191] Karin Stacke et al. “Measuring Domain Shift for Deep Learning in Histopathology”. en. In: *IEEE J Biomed Health Inform* 25.2 (Feb. 2021), pp. 325–336.

[192] Ming Y Lu et al. “Data-efficient and weakly supervised computational pathology on whole-slide images”. en. In: *Nat Biomed Eng* 5.6 (June 2021), pp. 555–570.

[193] Homeira Vafaei et al. “Association of placental chorangiosis with pregnancy complication and prenatal outcome: a case-control study”. en. In: *BMC Pregnancy Childbirth* 21.1 (Jan. 2021), p. 99.

[194] Raymond W Redline, Christina Bagby, and Sanjita Ravishankar. “Hypervascularity”. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 163–172.

[195] Erica Schollenberg et al. “Placenta and Pregnancy-Related Diseases”. In: *Gynecologic and Obstetric Pathology, Volume 2*. Ed. by Wenxin Zheng et al. Singapore: Springer Singapore, 2019, pp. 493–539.

[196] Claudia Vanea et al. “A New Graph Node Classification Benchmark: Learning Structure from Histology Cell Graphs”. In: New Frontiers in Graph Learning at NeurIPS, Nov. 2022.

[197] Philip J Katzman, Linda M Ernst, and Irene B Scheimberg. “Massive Perivillous Fibrinoid Deposition and Maternal Floor Infarct”. en. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 77–82.

[198] Raymond W Redline and Amy Pappin. “Fetal thrombotic vasculopathy: The clinical significance of extensive avascular villi”. en. In: *Hum. Pathol.* 26.1 (Jan. 1995), pp. 80–85.

[199] Beverly Barton Rogers et al. “Avascular Villi, Increased Syncytial Knots, and Hypervascular Villi Are Associated with Pregnancies Complicated by Factor V Leiden Mutation”. In: *Pediatr. Dev. Pathol.* 13.5 (2010), pp. 341–347.

[200] Cigdem Gunduz, Bülent Yener, and S Humayun Gultekin. “The cell graphs of cancer”. eng. In: *Bioinformatics* 20 Suppl 1 (Aug. 2004), pp. i145–151.

[201] Bülent Yener. “Cell-Graphs: Image-Driven Modeling of Structure-Function Relationship”. en. In: *Commun. ACM* 60.1 (2017), pp. 74–84.

[202] Cigdem Demir, S Humayun Gultekin, and Bülent Yener. “Augmented cell-graphs for automated cancer diagnosis”. en. In: *Bioinformatics* 21 Suppl 2 (Sept. 2005), pp. ii7–12.

[203] Cemal Cagatay Bilgin et al. “ECM-Aware Cell-Graph Mining for Bone Tissue Modeling and Classification”. en. In: *Data Min. Knowl. Discov.* 20.3 (Oct. 2009), pp. 416–438.

[204] Cemal Cagatay Bilgin et al. “Multiscale feature analysis of salivary gland branching morphogenesis”. en. In: *PLoS One* 7.3 (Mar. 2012), e32906.

[205] Korsuk Sirinukunwattana et al. “Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer”. en. In: *Sci. Rep.* 8.1 (Sept. 2018), p. 13692.

[206] Sajid Javed et al. “Cellular community detection for tissue phenotyping in colorectal cancer histology images”. en. In: *Med. Image Anal.* 63 (July 2020), p. 101696.

[207] Junbum Kim et al. “Unsupervised discovery of tissue architecture in multiplexed imaging”. en. In: *Nat. Methods* (Oct. 2022), pp. 1–9.

[208] Giovanni Palla et al. “Squidpy: a scalable framework for spatial omics analysis”. en. In: *Nat. Methods* 19.2 (Feb. 2022), pp. 171–178.

[209] Jiachen Li et al. “Cell clustering for spatial transcriptomics data with graph neural networks”. en. In: *Nat Comput Sci* 2.6 (June 2022), pp. 399–408.

[210] Shobana V Stassen et al. “PARC: ultrafast and accurate clustering of phenotypic data of millions of single cells”. In: *Bioinformatics* 36.9 (May 2020), pp. 2778–2786.

[211] Maria Brbić et al. “Annotation of spatially resolved single-cell data with STELLAR”. en. In: *Nat. Methods* (Oct. 2022), pp. 1–8.

[212] Richard J Chen et al. “Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis”. eng. In: *IEEE Trans. Med. Imaging* PP (Sept. 2020).

[213] Wenqi Lu et al. “Capturing Cellular Topology in Multi-Gigapixel Pathology Images”. en. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* Seattle, WA, USA: IEEE, 2020, pp. 1049–1058.

[214] Yanning Zhou et al. “CGC-Net: Cell Graph Convolutional Network for Grading of Colorectal Cancer Histology Images”. en. In: *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).* Seoul, Korea (South): IEEE, 2019, pp. 388–398.

[215] Helen Theissen et al. “Multi-Scale Graphical Representation of Cell Environment”. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC).* IEEE, July 2022, pp. 3522–3525.

[216] Jiangbo Shi et al. “A Structure-Aware Hierarchical Graph-Based Multiple Instance Learning Framework for pT Staging in Histopathological Image”. en. In: *IEEE Trans. Med. Imaging* 42.10 (Oct. 2023), pp. 3000–3011.

[217] Guillaume Jaume et al. *Towards Explainable Graph Representations in Digital Pathology.* July 2020.

[218] Pushpak Pati et al. “Hierarchical graph representations in digital pathology”. en. In: *Med. Image Anal.* 75 (Jan. 2022), p. 102264.

[219] Yuxuan Hu et al. “Unsupervised and supervised discovery of tissue cellular neighborhoods from cell phenotypes”. en. In: *Nat. Methods* (Jan. 2024), pp. 1–12.

[220] Jingwen Wang et al. “Weakly Supervised Prostate Tma Classification Via Graph Convolutional Networks”. English (US). In: *ISBI 2020 - 2020 IEEE International Symposium on Biomedical Imaging.* IEEE Computer Society, Apr. 2020, pp. 239–243.

[221] Richard J Chen et al. *Whole Slide Images are 2D Point Clouds: Context-Aware Survival Prediction using Patch-based Graph Convolutional Networks.* July 2021.

[222] Wenqi Lu et al. “SlideGraph+: Whole slide image level graphs to predict HER2 status in breast cancer”. In: *Med. Image Anal.* 80 (Aug. 2022), p. 102486.

[223] Yongju Lee et al. “Derivation of prognostic contextual histopathological features from whole-slide images of tumours via graph deep learning”. en. In: *Nat Biomed Eng* (Aug. 2022).

[224] Yi Zheng et al. “A Graph-Transformer for Whole Slide Image Classification”. en. In: *IEEE Trans. Med. Imaging* 41.11 (2022), pp. 3003–3015.

[225] Yu Zhao et al. “Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning With Deep Graph Convolution”. en. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020, pp. 4836–4845.

[226] Neda Zamanitajeddin et al. “Social network analysis of cell networks improves deep learning for prediction of molecular pathways and key mutations in colorectal cancer”. en. In: *Med. Image Anal.* 93 (Apr. 2024), p. 103071.

[227] Linda Studer et al. “Classification of Intestinal Gland Cell-Graphs Using Graph Neural Networks”. In: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, Jan. 2021, pp. 3636–3643.

[228] Xiangyan Meng and Tonghui Zou. “Clinical applications of graph neural networks in computational histopathology: A review”. en. In: *Comput. Biol. Med.* 164 (Sept. 2023), p. 107201.

[229] Bin Li, Yin Li, and Kevin W Eliceiri. “Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning”. In: (Nov. 2020). arXiv: `2011.08939 [cs.CV]`.

[230] Taimoor Shakeel Sheikh, Yonghee Lee, and Migyung Cho. “Histopathological Classification of Breast Cancer Images Using a Multi-Scale Input and Multi-Feature Network”. en. In: *Cancers* 12.8 (July 2020).

[231] Mostafa Jahanifar et al. “Robust Interactive Semantic Segmentation of Pathology Images with Minimal User Input”. en. In: *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Montreal, BC, Canada: IEEE, 2021, pp. 674–683.

[232] Simon Graham et al. “MILD-Net: Minimal information loss dilated network for gland instance segmentation in colon histology images”. en. In: *Med. Image Anal.* 52 (Feb. 2019), pp. 199–211.

[233] Mahendra Khened et al. “A generalized deep learning framework for whole-slide image segmentation and analysis”. en. In: *Sci. Rep.* 11.1 (June 2021), p. 11579.

[234] Ashley L Kiemen et al. “CODA: quantitative 3D reconstruction of large tissues at cellular resolution”. en. In: *Nat. Methods* 19.11 (Nov. 2022), pp. 1490–1499.

[235] Johnathan Pocock et al. “TIAToolbox as an end-to-end library for advanced tissue image analytics”. en. In: *Commun. Med.* 2 (Sept. 2022), p. 120.

[236] Babak Ehteshami Bejnordi et al. “Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies”. In: *Mod. Pathol.* 31.10 (Oct. 2018), pp. 1502–1512.

[237] Yan Xu et al. “Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features”. In: *BMC Bioinformatics* 18.1 (May 2017), p. 281.

[238] Yan Xu et al. “Weakly supervised histopathology cancer image segmentation and classification”. en. In: *Med. Image Anal.* 18.3 (Apr. 2014), pp. 591–604.

[239] Debora Kidron et al. “Automated image analysis of placental villi and syncytial knots in histological sections”. In: *Placenta* 53 (May 2017), pp. 113–118.

[240] Rashmi Mukherjee. “Morphometric Evaluation of Preeclamptic Placenta Using Light Microscopic Images”. In: *Biomed Res. Int.* 2014 (2014), p. 293690.

[241] Zaneta Swiderska-Chadaj et al. “Image processing methods for the structural detection and gradation of placental villi”. en. In: *Comput. Biol. Med.* 100 (Sept. 2018), pp. 259–269.

[242] Alexander Maly et al. “Histomorphometric study of placental villi vascular volume in toxemia and diabetes”. en. In: *Hum. Pathol.* 36.10 (Oct. 2005), pp. 1074–1079.

[243] Stefano Marletta et al. “Application of Digital Imaging and Artificial Intelligence to Pathology of the Placenta”. eng. In: *Pediatr. Dev. Pathol.* 26.1 (2023), pp. 5–12.

[244] D Eppstein, M S Paterson, and F F Yao. “On Nearest-Neighbor Graphs”. en. In: *Discrete Comput. Geom.* 17.3 (Apr. 1997), pp. 263–282.

[245] Leonidas J Guibas, Donald E Knuth, and Micha Sharir. “Randomized incremental construction of Delaunay and Voronoi diagrams”. en. In: *Algorithmica* 7.1 (June 1992), pp. 381–413.

[246] Hongbin Pei et al. “Geom-GCN: Geometric Graph Convolutional Networks”. In: (Feb. 2020). arXiv: `2002.05287 [cs.LG]`.

[247] Raymond W Redline and Sanjita Ravishankar. “Fetal vascular malperfusion, an update”. en. In: *APMIS* 126.7 (2018), pp. 561–569.

[248] Tong Zhao et al. *Graph Data Augmentation for Graph Machine Learning: A Survey*. Jan. 2023.

[249] Kaize Ding et al. “Data Augmentation for Deep Graph Learning: A Survey”. In: *SIGKDD Explor. Newsl.* 24.2 (Dec. 2022), pp. 61–77.

[250] Jonathan Campbell et al. “Enhancing Cross-Institute Generalisation of GNNs in Histopathology through Multiple Embedding Graph Augmentation (MEGA)”.

[251] Oleksandr Shchur et al. “Pitfalls of Graph Neural Network Evaluation”. In: *arXiv:1811.05868 [cs, stat]* (June 2019).

[252] Zhilin Yang, William W Cohen, and Ruslan Salakhutdinov. *Revisiting Semi-Supervised Learning with Graph Embeddings*. May 2016.

[253] Andrew Carlson et al. “Toward an Architecture for Never-Ending Language Learning”. en. In: *AAAI Press* (2010), p. 8.

[254] Petar Veličković et al. *Graph Attention Networks*. Feb. 2018.

[255] Marinka Zitnik and Jure Leskovec. “Predicting multicellular function through multi-layer tissue networks”. In: *Bioinformatics* 33.14 (July 2017), pp. i190–i198.

[256] Felix Wu et al. *Simplifying Graph Convolutional Networks*. June 2019.

[257] Petar Veličković et al. *Deep Graph Infomax.* Dec. 2018.

[258] Damian Szklarczyk et al. “STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets”. In: *Nucleic Acids Res.* 47.D1 (Jan. 2019), pp. D607–D613.

[259] Hanqing Zeng et al. *Decoupling the Depth and Scope of Graph Neural Networks.* Jan. 2022.

[260] Fabrizio Frasca et al. “SIGN: Scalable Inception Graph Neural Networks”. en. In: *ICML 2020.* Apr. 2020.

[261] Shaked Brody, Uri Alon, and Eran Yahav. *How Attentive are Graph Attention Networks?* Jan. 2022.

[262] Korsuk Sirinukunwattana et al. “Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images”. en. In: *IEEE Trans. Med. Imaging* 35.5 (May 2016), pp. 1196–1206.

[263] “Partitioning around medoids (program PAM)”. In: *Finding Groups in Data.* Hoboken, NJ, USA: John Wiley & Sons, Inc., 2008, pp. 68–125.

[264] J MacQueen. “Some methods for classification and analysis of multivariate observations”. en. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics.* Vol. 5.1. University of California Press, Jan. 1967, pp. 281–298.

[265] Terry K Morgan. “Frontiers in Placental Pathology”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 379–382.

[266] T Yee Khong et al. “Sampling and Definitions of Placental Lesions: Amsterdam Placental Workshop Group Consensus Statement”. In: *Arch. Pathol. Lab. Med.* 140.7 (May 2016), pp. 698–713.

[267] Leigh A Taylor et al. *How often is the placenta included in human pregnancy research? A rapid systematic review of the literature.* en. Tech. rep. 5:38. Gates Open Research, Mar. 2021.

[268] M Castellucci et al. “The development of the human placental villous tree”. eng. In: *Anat. Embryol.* 181.2 (1990), pp. 117–128.

[269] Kurt Benirschke, Graham J Burton, and Rebecca N Baergen. “Basic Structure of the Villous Trees”. In: *Pathology of the Human Placenta.* Ed. by Kurt Benirschke, Graham J Burton, and Rebecca N Baergen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 55–100.

[270] Philip J Katzman. “Chorionic Plate Chronic Inflammatory Lesions Including Eosinophilic/T-Cell Chorionic Vasculitis”. en. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 109–113.

[271] Chong Jai Kim and Jung-Sun Kim. “Chronic Villitis”. In: *Pathology of the Placenta: A Practical Guide.* Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 195–206.

[272] Labelbox. *Labelbox.* `https://labelbox.com`. 2022.

[273] Mary L McHugh. “Interrater reliability: the kappa statistic”. In: *Biochem. Med.* 22.3 (Oct. 2012), pp. 276–282.

[274] L Glasser et al. “A comprehensive study of umbilical cord blood cell developmental changes and reference ranges by gestation, gender and mode of delivery”. en. In: *J. Perinatol.* 35.7 (July 2015), pp. 469–475.

[275] Leland McInnes, John Healy, and James Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. en. In: (Feb. 2018).

[276] Sanne J Gordijn et al. “Constellations of Pathology in the Placenta and How They Relate to Clinical Conditions”. en. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 361–369.

[277] Janet M Catov et al. “Maternal Vascular Lesions in the Placenta Predict Vascular Impairments a Decade After Delivery”. en. In: *Hypertension* 79.2 (Feb. 2022), pp. 424–434.

[278] Linda M Ernst. “Maternal vascular malperfusion of the placental bed”. en. In: *APMIS* 126.7 (July 2018), pp. 551–560.

[279] Hatem A Mousa and Zarko Alfirevic1. “Do placental lesions reflect thrombophilia state in women with adverse pregnancy outcome?” In: *Hum. Reprod.* 15.8 (Aug. 2000), pp. 1830–1833.

[280] D M O Becroft, J M D Thompson, and E A Mitchell. “The Epidemiology of Placental Infarction at Term”. en. In: *Placenta* 23.4 (Apr. 2002), pp. 343–351.

[281] Louise Devisme et al. “Perinatal outcome of placental massive perivillous fibrin deposition: a case-control study”. en. In: *Prenat. Diagn.* 37.4 (Apr. 2017), pp. 323–328.

[282] Ira Adams-Chapman et al. “Maternal floor infarction of the placenta: association with central nervous system injury and adverse neurodevelopmental outcome”. en. In: *J. Perinatol.* 22.3 (2002), pp. 236–241.

[283] Eric K Morgen, Brendan Fitzgerald, and Sarah Keating. “Intervillous Thrombosis”. en. In: *Pathology of the Placenta: A Practical Guide*. Ed. by T Yee Khong et al. Cham: Springer International Publishing, 2019, pp. 67–76.

[284] Rebecca N Baergen. “Miscellaneous Placental Lesions”. In: *Manual of Pathology of the Human Placenta: Second Edition*. Ed. by Rebecca N Baergen. Boston, MA: Springer US, 2011, pp. 355–377.

[285] Linda M Ernst et al. “Placental Pathologic Associations With Morbidly Adherent Placenta: Potential Insights Into Pathogenesis”. en. In: *Pediatr. Dev. Pathol.* 20.5 (Mar. 2017), pp. 387–393.

[286] Afsoon Khodaee et al. “Automatic Placental Distal Villous Hypoplasia Scoring using a Deep Convolutional Neural Network Regression Model”. In: *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, May 2022, pp. 1–5.

[287] Thomas N Kipf and Max Welling. “Variational Graph Auto-Encoders”. en. In: (Nov. 2016).

[288] Shirui Pan et al. “Adversarially Regularized Graph Autoencoder for Graph Embedding”. en. In: (Feb. 2018).

[289] Zhenyu Hou et al. *GraphMAE: Self-Supervised Masked Graph Autoencoders.* July 2022.

[290] Zhenyu Hou et al. *GraphMAE2: A Decoding-Enhanced Masked Self-Supervised Graph Learner.* Apr. 2023.

[291] Chun Wang et al. “MGAE: Marginalized Graph Autoencoder for Graph Clustering”. en. In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* Singapore Singapore: ACM, Nov. 2017, pp. 889–898.

[292] Hongyang Gao and Shuiwang Ji. “Graph U-Nets”. In: *arXiv:1905.05178 [cs, stat]* (May 2019).

[293] Qichao Liu et al. “Multilevel Superpixel Structured Graph U-Nets for Hyperspectral Image Classification”. In: *IEEE Trans. Geosci. Remote Sens.* 60 (2022), pp. 1–15.

[294] Guillaume Salha et al. *FastGAE: Scalable Graph Autoencoders with Stochastic Subgraph Decoding.* Apr. 2021.

[295] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. “Self-Attention Graph Pooling”. In: *arXiv:1904.08082 [cs, stat]* (June 2019).

[296] Boris Knyazev, Graham W Taylor, and Mohamed R Amer. “Understanding Attention and Generalization in Graph Neural Networks”. In: *arXiv:1905.02850 [cs, stat]* (Oct. 2019).

[297] Ekagra Ranjan, Soumya Sanyal, and Partha Pratim Talukdar. “ASAP: Adaptive Structure Aware Pooling for Learning Hierarchical Graph Representations”. In: *arXiv:1911.07979 [cs, stat]* (Feb. 2020).

[298] Rex Ying et al. “Hierarchical Graph Representation Learning with Differentiable Pooling”. en. In: (June 2018).

[299] Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. “Spectral Clustering with Graph Neural Networks for Graph Pooling”. In: (June 2019). arXiv: `1907.00481 [cs.LG]`.

[300] Charles R Qi et al. “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”. In: (June 2017). arXiv: `1706.02413 [cs.CV]`.

[301] Frederik Diehl et al. “Towards graph pooling by edge contraction”. In: *ICML 2019 Workshop on Learning and Reasoning with Graph-Structured Data* (2019).

[302] Frederik Diehl. “Edge Contraction Pooling for Graph Neural Networks”. In: (May 2019). arXiv: `1905.10990 [cs.LG]`.

[303] Hengshuang Zhao et al. “Point Transformer”. In: (Dec. 2020). arXiv: `2012.09164 [cs.CV]`.

[304] Sergey Ioffe and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: (Feb. 2015). arXiv: `1502.03167 [cs.LG]`.

[305] Diederik P Kingma and Max Welling. “An Introduction to Variational Autoencoders”. In: (June 2019). arXiv: `1906.02691 [cs.LG]`.

[306] Keyulu Xu et al. “How Powerful are Graph Neural Networks?” In: (Oct. 2018). arXiv: `1810.00826 [cs.LG]`.

[307] Devin Kreuzer et al. "Rethinking Graph Transformers with Spectral Attention". In: (June 2021). arXiv: `2106.03893 [cs.LG]`.

[308] Chen Cai et al. "On the Connection Between MPNN and Graph Transformer". In: (Jan. 2023). arXiv: `2301.11956 [cs.LG]`.

[309] Jan Tönshoff et al. "Where Did the Gap Go? Reassessing the Long-Range Graph Benchmark". In: (Sept. 2023). arXiv: `2309.00367 [cs.LG]`.

[310] Francesco Cisternino et al. *Self-supervised learning for characterising histomorphological diversity and spatial RNA expression prediction across 23 human tissue types.* en. Aug. 2023.

[311] Emiel Hoogeboom et al. "Equivariant Diffusion for Molecule Generation in 3D". In: (Mar. 2022). arXiv: `2203.17003 [cs.LG]`.

[312] Clement Vignac et al. "DiGress: Discrete Denoising diffusion for graph generation". In: (Sept. 2022). arXiv: `2209.14734 [cs.LG]`.

[313] Ilia Igashov et al. "Equivariant 3D-Conditional Diffusion Models for Molecular Linker Design". In: (Oct. 2022). arXiv: `2210.05274 [cs.LG]`.

[314] Konpat Preechakul et al. "Diffusion Autoencoders: Toward a Meaningful and Decodable Representation". In: (Nov. 2021). arXiv: `2111.15640 [cs.CV]`.

[315] Weilai Xiang et al. "Denoising Diffusion Autoencoders are Unified Self-supervised Learners". In: (Mar. 2023). arXiv: `2303.09769 [cs.CV]`.

[316] *Moorfields and DeepMind: bringing A.I. closer to the eye clinic.* en. `https://moorfieldsbrc.nihr.ac.uk/case-study/deepmind-research-report/`. Accessed: 2024-4-14. June 2023.

[317] Jeffrey De Fauw et al. "Clinically applicable deep learning for diagnosis and referral in retinal disease". en. In: *Nat. Med.* 24.9 (Sept. 2018), pp. 1342–1350.

[318] Yukun Zhou et al. "A foundation model for generalizable disease detection from retinal images". en. In: *Nature* 622.7981 (Oct. 2023), pp. 156–163.

[319] Nisha Sharma et al. "Multi-vendor evaluation of artificial intelligence as an independent reader for double reading in breast cancer screening on 275,900 mammograms". en. In: *BMC Cancer* 23.1 (May 2023), p. 460.

[320] Jong Seok Ahn et al. "Association of Artificial Intelligence-Aided Chest Radiograph Interpretation With Reader Performance and Efficiency". en. In: *JAMA Netw Open* 5.8 (Aug. 2022), e2229289.

[321] Pranav Rajpurkar et al. "CheXaid: deep learning assistance for physician diagnosis of tuberculosis using chest x-rays in patients with HIV". en. In: *NPJ Digit Med* 3 (Sept. 2020), p. 115.

[322] Susanne Gaube et al. "Non-task expert physicians benefit from correct explainable AI advice when reviewing X-rays". en. In: *Sci. Rep.* 13.1 (Jan. 2023), p. 1383.

[323] Alexander Hodge et al. "Human Amnion Epithelial Cells Produce Soluble Factors that Enhance Liver Repair by Reducing Fibrosis While Maintaining Regeneration in a Model of Chronic Liver Injury". en. In: *Cell Transplant.* 29 (2020), p. 963689720950221.

[324] Gita Pratama et al. “Changes in culture expanded human amniotic epithelial cells: implications for potential therapeutic applications”. en. In: *PLoS One* 6.11 (Nov. 2011), e26136.

[325] Chen Qiu et al. “Human Amniotic Epithelial Stem Cells: A Promising Seed Cell for Clinical Applications”. en. In: *Int. J. Mol. Sci.* 21.20 (Oct. 2020).

[326] Yuhan Du et al. “The Role of XAI in Advice-Taking from a Clinical Decision Support System: A Comparative User Study of Feature Contribution-Based and Example-Based Explanations”. en. In: *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* 12.20 (Oct. 2022), p. 10323.

[327] Jennifer S N Tang et al. “Impact of Different Artificial Intelligence User Interfaces on Lung Nodule and Mass Detection on Chest Radiographs”. en. In: *Radiol Artif Intell* 5.3 (May 2023), e220079.

[328] Clarisse F de Vries et al. “Impact of Different Mammography Systems on Artificial Intelligence Performance in Breast Cancer Screening”. en. In: *Radiol Artif Intell* 5.3 (May 2023), e220146.

[329] Weixin Liang et al. “Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews”. In: (Mar. 2024). arXiv: `2403.07183 [cs.CL]`.

[330] Cheng Wang. “Calibration in Deep Learning: A Survey of the State-of-the-Art”. In: (Aug. 2023). arXiv: `2308.01222 [cs.LG]`.

[331] Jane Wakefield. “Fitbit faces anger for setting limits on women’s periods”. en. In: *BBC* (Aug. 2018).

[332] Yuejiang Yu et al. “Do Deep Learning Models Really Outperform Traditional Approaches in Molecular Docking?” In: (Feb. 2023). arXiv: `2302.07134 [q-bio.BM]`.

[333] Gengmo Zhou et al. “Do Deep Learning Methods Really Perform Better in Molecular Conformation Generation?” In: (Feb. 2023). arXiv: `2302.07061 [cs.CE]`.